Theory and Applications of Electricity and Magnetism

THEORY AND APPLICATIONS OF

Electricity and Magnetism

By Charles A. Culver

Formerly, Head of the Department of Physics, Carleton College

Fellow, American Institute of Electrical Engineers

FIRST EDITION

SECOND IMPRESSION

NEW YORK AND LONDON

McGRAW HILL BOOK COMPANY, INC.

1947

THEORY AND APPLICATIONS
OF
ELECTRICITY AND MAGNETISM

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

The present volume is, in part, the outgrowth of a text published by the author in 1930. That book, bearing the title "Electricity and Magnetism," was designed to serve as the basis of a 3-hr semester course, following a course in general college physics. The original text went through several printings, but because of the exigencies of war, with the resulting destruction of the plates, the book is now out of print. Because of this fact, and because the field has greatly expanded, the author has been led to prepare a new and more comprehensive text.

The present offering is designed to present the material which would normally be covered in a 3-hr year course given to juniors and seniors in colleges and universities. The text is based on the assumption that the student has had a thorough course in general college physics and that he is also acquainted with differential and integral calculus. The material presented is intended to serve as a basis for advanced study in physics and chemistry and also to lay the foundation for courses in electrical engineering.

The text embodies the experience gained from many years of teaching the subject and from engineering practice. The constructive criticism of both teachers and students who used the original shorter text has also been an important factor in the preparation of the present volume. It is believed that the order of presentation will be found to be logical and adapted to students' needs.

The electron theory of matter has been adopted throughout the text. Such a procedure raises the question as to what convention should be followed in regard to the direction of the electric current. It is fairly well agreed that the time has come when formal steps should be taken to abandon the ancient convention that the current proceeds from a point of high potential to a point of lower potential. However, the older convention is so thoroughly incorporated in the literature of the subject that it has been felt that undue confusion would result if a single text were to adopt the newer (and correct) convention. Accordingly, the author has, in the main, followed the established custom. Often the direction of both the electron and the conventional current has been indicated. In discussing electrolytic and electronic phenomena the direction of electrons is in all cases indicated.

It is to be hoped that, in the near future, the American Institute of Electrical Engineers and the Institute of Radio Engineers will jointly take formal action in this matter, to the end that our conventions may conform to known facts, rather than tradition.

The terminology and symbols are in accord with the recommendations of the AIEE.

In the discussion of principles and laws an effort has been made to lead the student to see how analytical tools, in the form of working equations, are developed and applied. In that connection special attention has been given to a discussion of the units involved. Certain analytical problems are left for the student to solve.

In a few instances a concept has been referred to before it has been discussed in detail. In a subject as broad as modern electricity and magnetism, this is inevitable. However, in such cases reference to the place in the text where that particular subject is discussed is always given.

The problems are more or less graded as to difficulty, and the data involved are taken largely from actual cases.

References to specialized texts and original papers will be found embodied in the text itself rather than in footnotes. It has been the author's experience that students pay little attention to footnotes or to references placed at the end of chapters. Students should be encouraged to examine original sources.

It is the hope of the author that the present volume will be as favorably received as its predecessor and that it may be found to serve as a thorough foundation for advanced study in this field. Information concerning any errors which may be noted will be gratefully received.

CHARLES A. CULVER

Northfield, Minn.
July, 1947

CONTENTS

CHAPTER I

INTRODUCTORY

1. The Constitution of Matter. Any discussion of the theories, concepts, laws, and applications of electricity and magnetism necessarily involves an examination of our understanding of the nature of matter. At the present time the picture may be sketched as follows.

All atoms are considered to be made up of certain electrical entities, these entities being associated in various ways, and subject to certain laws. In certain respects the most important of these entities is the **electron,** the smallest quantity of negative electricity yet isolated. As we shall see later, the orderly migration of electrons constitutes what we commonly speak of as an electric current. An atom appears to consist of a relatively massive central assembly of entities called the nucleus, about which revolve definite configurations of electrons. It is thought that each orbital electron spins about a symmetrical axis.

A second elementary electrical entity, the **positron,** having the same mass as the electron is known to exist; though its existence as an independent entity is more or less transient. This entity will not concern us greatly.

The basic positive electrical charge is the **proton,** having a charge equal to that of the electron and a mass approximately 1,838 times that of the electron. The proton constitutes the nucleus of the ordinary hydrogen atom.

There is also evidence tending to indicate the existence of an entity known as the **mesotron.** This "particle" appears to have a mass of the order of 200 times that of the electron. Some mesotrons have been observed which exhibit a positive charge; others show a negative charge. These entities are transient and will enter but little into the discussions which follow.

A fifth entity is known as the **neutron.** This "particle" is considered to be made up of a proton closely associated with an electron. As its name implies, it does not manifest a charge.

Still another "particle" is designated as a **deuteron.** This entity consists of a neutron and a proton. It exhibits a positive charge, and is in fact the nucleus of the heavy hydrogen atom. It has twice the mass of the proton.

An entity which will enter into our discussions in a later chapter is referred to as the **alpha particle.** This particle consists of a fairly closely

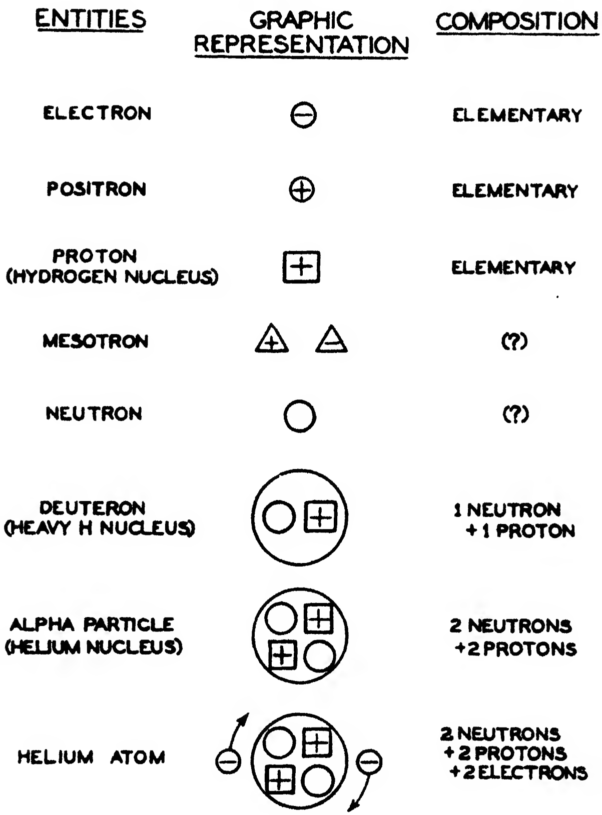
associated group of two neutrons and two protons. It exhibits two positive units of charge and has a mass four times that of the proton.

The manner in which these various entities are associated in an atom is illustrated in the structure of the helium atom. The nucleus of an atom of helium gas consists of two neutrons and two protons. Two orbital electrons complete the structure (see last item in the following chart). The nucleus of the helium atom constitutes the α-particle, as indicated above. The orbital electrons of all atoms move about the nucleus in definite paths; the orbits are probably not coplanar, and have different radii. For reasons which we shall examine later, a given orbital electron may shift its orbit to one having a smaller or a greater radius. Some of the orbital electrons of certain atoms may be detached from the atomic structure of which they form a part. The means whereby this may be accomplished will form the subject matter of a later chapter. In the case of metals at least some of the orbital electrons appear to move more or less freely from atom to atom, though at any given instant each atom has its normal complement of electrons. These "meandering" electrons are referred to as **free** electrons. We shall find that, by suitable means, it is possible to cause these free electrons to move in an orderly fashion, thus establishing what has already been referred to as an electric current. Bodies that exhibit this electronic behavior are classified as **conductors.** There are, however, bodies in which little if any interatomic movement obtains; such bodies are referred to as **nonconductors** or **insulators.** The chart that follows may serve to show, in a diagrammatic way, the facts briefly referred to above. The foregoing outline, together with the following chart, will serve as a preliminary basis for our discussion of charges at rest (electrostatics) and charges in motion (electrodynamics). Later we shall examine the subject of the electrical nature of matter in greater detail.

2. A Charge. It is important to arrive at a clear understanding of the terms "a charge" and "a charged body." What do these expressions mean? On the basis of the electron theory of matter, neutral atoms are those in which just enough electrons are present to neutralize the positive charges (protons) in the nucleus. If, by any means, one or more of the electrons of an atom be removed, the remainder of the atomic structure will manifest a positive charge, and the body of which it forms a part will exhibit a positive charge. If, for example, one of the orbital electrons of a helium atom were to be removed, the remainder of the atoms would show a positive charge. For instance, if a piece of silk cloth be brought into intimate physical contact with a glass rod and then these two bodies are separated, it will be found that the silk has gained a number of electrons and the glass has lost **an equal number.** Thus the

glass becomes positively charged and the silk negatively charged. Electrification is thus seen to consist of the transfer of electrons to or from a body. A charged body thus may be said to be one that has a shortage, or an excess, of electrons. It will thus be seen that we are to be concerned, chiefly, with electrons. It will therefore be convenient, in the discussions which are to follow, to think of electrons as discrete entities.

NUCLEAR CHART



3. Units of Charge. As we proceed, we shall have occasion to deal with definite numbers of electrons, *i.e.*, with specific quantities of charge. It becomes necessary, therefore, to establish a unit of electrical quantity.

As already indicated, the smallest definite quantity of electricity (or charge) that has been isolated is the electron, and all larger quantities are found to be exact multiples of this fundamental unit. Since the electron is an exceedingly small quantity of electricity, it is customary to

use a larger unit of quantity. Since, as we shall see later, electrical charges produce certain definite mechanical force actions, it is convenient to define unit charge in terms of the fundamental unit of force. Stated thus, it may be said that **unit quantity of electricity is that quantity which when concentrated at a point in a vacuum at a distance of 1 cm from an equal and similar quantity of charge will be repelled by a force of 1 dyne.** This is an arbitrary, but convenient, quantity in which to express the magnitude of electrical charges. It will be noted that we here revert to the basic unit in mechanics as a starting point from which to build up a system of electrical units. This intimate relation between the science of mechanics and the domain of electricity will become more apparent as we proceed.

The unit above defined is known as the **electrostatic unit of charge** and is commonly abbreviated by the letters esu. In terms of this unit, the electron has a value of $(4.8029 \pm 0.0005) \times 10^{-10}$ esu. For most purposes of ordinary computation 4.8×10^{-10} may be taken as a working value. This electronic charge value is commonly designated by the letter e. The electronic charge e is one of the basic quantities in science and is comparable in importance with such quantities as g, J, h, and c. The student should fix the magnitude of this quantity clearly in mind. While the electrostatic unit of quantity is very much larger than the charge represented by the electron (there are more than 2.08×10^9—over 2 billion—electrons in 1 esu), it is, however, inconveniently small when dealing with practical quantities of electricity. Hence a multiple of the electrostatic unit has been agreed upon; it is known as the **coulomb** and is equivalent to 3×10^9 esu. On this basis the electron has a value of 1.601×10^{-19} coulomb. The reason for taking the particular number of esu, as indicated above, to serve as a practical unit of quantity (the coulomb) will become apparent in later discussions.

The electrostatic unit of quantity is frequently referred to by the name **statcoulomb.** Statcoulomb and electrostatic unit are therefore synonymous.

For convenience, positive charges are designated by a + sign, and negative charges by a − sign.

CHAPTER II

THE ELECTRIC FIELD

4. First Law of Electrostatics. Du Fay, an early French investigator, found that **unlike charges exert an attractive force for one another and like charges a repellent force.** This may be thought of as the **first law of electrostatics.** No satisfactory explanation of this force action has as yet been arrived at by the scientific world. Apparently Faraday could not accept the idea of "action at a distance." To him, and to others, it seemed more reasonable to assume some sort of a universal medium, and that electrostatic stresses developed strains in this hypothetical medium. On this basis Faraday introduced certain conventions that, it may be said in passing, still serve a useful purpose. He thought of **lines of force** having their origin on positive charges and terminating on negative charges. Faraday considered these lines to be under tension longitudinally and subject to mutual repulsion laterally. Today we utilize these hypothetical lines to represent direction of force action, and the number of the lines per unit area to represent what we shall presently define as field strength. Whether we postulate the existence of a universal medium, or proceed on the basis of "action at a distance," we do know that the space in the vicinity of a charge exhibits certain special characteristics, and that these properties can be graphically represented by the lines-of-force convention. **A region in which an exploring (test) charge experiences an electrostatic force action is said to constitute an electrostatic field.** A graphic representation, in one plane, of such a field in the region of two equal charges is shown in Fig. 1.

There is no such thing as a completely isolated charge. The lines of force in Fig. 1*b* terminate somewhere on an equal and opposite charge. If, however, a given charge is **far removed** from its complementary charge we sometimes speak of it as an "isolated charge." The arrows indicate the direction in which a positive test charge would tend to move. In making use of these conventions, however, the student should keep clearly in mind that **the lines have no physical existence;** they are used solely to assist in setting up a graphic representation of a field of force.

5. Second Law of Electrostatics. In 1784, Coulomb reported the results of an important experimental investigation which he had carried out to determine the relation between the magnitude of reacting charges and the resulting force. By means of a torsion balance, he compared the mechanical force between two small electrified bodies when separated

by different distances in air. While the experimental conditions were not entirely satisfactory, owing to the fact that the charges could not be concentrated at points, the data tended to show that an inverse square law obtained. The empirical evidence indicated that the force between two charges could be expressed by the relation

$$\text{Mechanical force} = F = \frac{qq'}{d^2} \tag{1}$$

when q and q' represent the magnitudes of the two charges involved (defined as in Sec. 3), F the force in dynes, and d the distance (in cms)

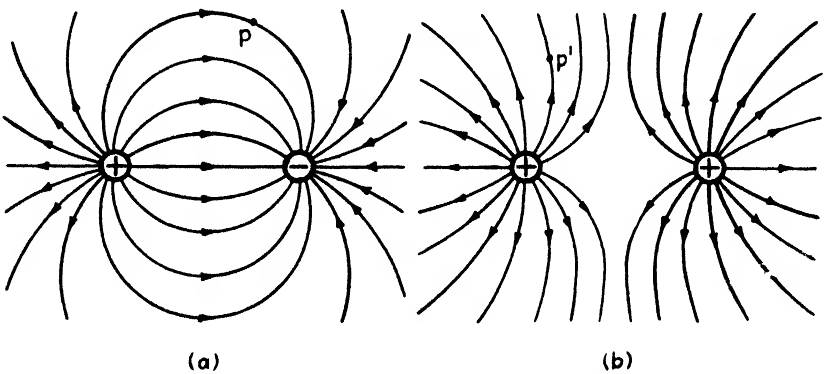


FIG. 1.—Electrostatic field in the region of adjacent like and unlike charges.

separating the two charges. This law holds for charges in free space, or for charges which form a part of what we speak of as a "material body." In the latter case, the body, of which they are a part, will experience the force. It has been shown by analytical methods that an inverse square law is the only relation which is consonant with the known physical phenomena. In fact, it has recently been shown[1] that the exponent of d in Eq. (1) is correct to within one part in a billion.

Subsequent to Coulomb's work, it was found by Cavendish and Faraday that the nature of the physical medium separating the two charges is also a factor in determining the magnitude of the mutual force exhibited by the charges. They found that the electromechanical force acting between charged bodies is, in general, **less** when a nonconductor (insulator) other than air separates the charges. For instance, the force is only about half as great when two charges are surrounded by some oil, such as kerosene, as when the intervening medium is air. The ratio of the force in the first case to the force in the second instance is called the **dielectric constant,** and is usually indicated by the letter K. For reasons that will

[1] PLIMPTON and LAWTON, *Phys. Rev.*, **50**, 1066 (1936).

appear later, Faraday used the term **specific inductive capacity,** to designate this factor. In view of the findings referred to above, it becomes necessary to introduce the factor K into the electrostatic force relation. The expression accordingly becomes

$$F = \frac{qq'}{Kd^2}. \tag{2}$$

It is now known that K is not unity except in a vacuum. The value for air is 1.00059. This value is, however, so nearly unity that it is commonly so considered in practice.

A practical problem will serve to illustrate the utility of the relation embodied in Eq. (2).

Problem. Assume two concentrated charges of +5 and −10 esu (statcoulombs) respectively to be 10 cm apart in air. What will be the resulting force? What would be the magnitude of the force if the same two charges were immersed in oil having a dielectric constant of 2.5?

Solution. Substituting in Eq. (2), we have

$$F = \frac{qq'}{Kd^2} = \frac{(+5)(-10)}{10^2} = -0.5 \text{ dyne.}$$

The negative sign indicates an attractive force. If the sign had turned out to be +, it would have indicated that the force was one of repulsion. In the second case,

$$F = \frac{(+5)(-10)}{2.5 \times 10^2} = -0.2 \text{ dyne.}$$

One additional word regarding the factor K: In order to separate two unlike charges force is required; work must be done. Since the force required depends upon the nature of the medium it must follow that **the energy involved in the process of separation of the charges resides in the dielectric** and not in the charges themselves. We shall return to this thought later.

6. The Charging of a Body. In any discussion of electric charges the question at once presents itself as to how one may bring about either a transient or a permanent displacement of electrons. As we shall see later, there are a number of methods whereby this end may be attained.

Perhaps the simplest, and possibly the first, method employed for the purpose of separating electrons from nuclei consists in bringing two dissimilar nonconductors, such as a piece of paper and a piece of woolen cloth, into intimate physical contact and then mechanically separating them. Upon separation the paper will manifest a charge opposite in character to that shown by the cloth. Why? The electrons of one of

the materials are probably held within the atomic structure more strongly than those in the atoms of the other material; hence, more electrons will be disengaged from one kind of material than from the other. The material whose atoms lose the greater number of electrons will thus manifest a positive charge, and the other material will show a negative charge because it will have gained the electrons lost by the other piece of material. Such a procedure is commonly referred to as **"charging by induction."** The term "induction" is unfortunate because it is later used in another sense. However, a careful reading of the following and subsequent statements will serve to avoid confusion.

The process of charging by induction may be readily understood by reference to the electrical situation as sketched in Fig. 2. In the series

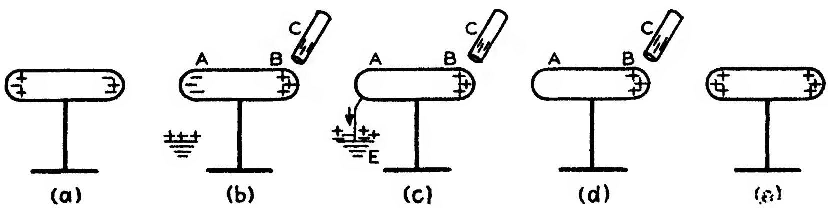


Fig. 2.—Charging by induction.

of sketches, (*a*) represents a conducting body attached to an insulating support. Electrically, it is neutral; the free electrons are so distributed that, at any given instant, each atom has its normal quota of electrons.

Suppose now that we bring a charged body *C*, such as a hard-rubber rod, near one end of the body, as shown in (*b*). Because of the laws of electrostatic force action, already referred to, at least some of the free electrons (indicated by the negative signs) will be repelled toward the end *A*, thus leaving the end on the right positively charged. The electrons at the nearest point on the earth will be repelled because of the electrons on *A*, thus leaving that part of the earth, or earth connection, positively charged, as shown in (*b*). If now one were to remove the negatively charged body *C*, it would be found that the electrons would at once redistribute themselves throughout the body, and the situation would revert to that pictured in (*a*). If, however, before removing *C* from the immediate vicinity of *B*, one connects *A* to the ground, the repelled electrons on *A* will be attracted by the positive charges at *E* and will therefore pass to earth, thus leaving the body with a permanent deficit of electrons. In other words, the body will be found to be **positively charged.** If *C* is removed, the entire surface of the original body will be found to be positively charged, as indicated in (*e*). If *C* had been positively charged, a corresponding process would have taken place;

in that case, negative charges (electrons) would have passed from the earth to A, thus leaving the body negatively charged. By means of a procedure essentially like the one outlined above, a body may be charged **inductively,** and the energy represented by the "induced" charge may be utilized at will. **Electricity has not been created;** rather have charges been more or less definitely **separated.** If the energy of a body, in the form of a charge, is utilized, the procedure outlined above could be repeated by the **expenditure of mechanical energy** (the moving of the rod); and thus mechanical energy can be converted into electrical energy.

There are several machines by which a "repeat" process may be continuously carried out and the stored charges thus made available for a particular use. The Wimshurst electrostatic "generator," which is described in elementary texts, is such a machine. A more modern organization is one known as the Van de Graaff "generator." For a description of this machine, the reader should consult an original article by Van de Graaff, Compton, and Van Atta, which appeared in the *Physical Review*, for Feb. 1, 1933.

7. Field Strength. In Sec. 4 reference was made to the electrostatic field and to the fact that such a region exhibits certain well-defined characteristics. It becomes necessary to have available some means whereby these characteristics may be described quantitatively. To accomplish this, recourse is had to the concept of **field strength,** or **field intensity.** If we assume that we have a given concentrated charge located at a certain point, and we place a unit test charge (say, positive) at a given point in the region of the assumed charge, we may ask ourselves this question: What electromechanical force will our test charge experience? If we should find that our unit test charge were subject to a force of 5 dynes, for example, we would say that the field strength is five units *i.e.*, 5 dynes per unit charge. To represent such a condition graphically, five lines would be drawn through an area of 1 cm², the area to include the point in question.

Our next question is: What quantitative relation exists between the field strength and the other factors involved?—viz., the magnitude of the charge and its distance from the point in question.

If in Eq. (2) we let q', for instance, represent the unit test charge, the relation becomes

$$F = \frac{q}{Kd^2} = \mathcal{E}, \tag{3}$$

where $\mathcal{E}$ represents the **field strength,** or **field intensity,** expressed in dynes per unit charge.

From the foregoing it will be evident that the total electromechanical force exerted upon any charge placed at a point in a field will be given by

the product of the field strength and the charge. This may be expressed by the simple relation

$$F = \mathcal{E} \times Q \tag{4}$$

If $\mathcal{E}$ is in esu (dynes per unit charge) and Q in esu (statcoulombs), F will be in dynes of force.

Problem. By way of illustration, let us assume a point 30 cm from a concentrated charge of 100 statcoulombs. What force will be experienced by a charge of 1.5×10^{-8} coulomb if placed at that point? The medium is air.

Solution. Making use of Eq. (3), we have

$$\mathcal{E} = \frac{q}{Kd^2} = \frac{100}{1 \times 30^2} = 0.111 \text{ dynes/unit charge.}$$

Substituting the above value for $\mathcal{E}$ in Eq. (4),

$$F = \mathcal{E} \times Q = 0.111 \times 1.5 \times 10^{-8} \times 3 \times 10^9 = 5 \text{ dynes.}$$

In considering problems involving the concept of field strength, it should be borne in mind that **we are dealing with a vector quantity.** Field intensity is expressed in terms of **force** and hence may be dealt with according to the usual vectorial procedure. Further, if we are dealing with a case that involves the field strength at a given point, due to the presence of several charges, each charge produces its own effect, regardless of the presence of the other charges.

8. Electrostatic Induction. In Sec. 7 it was indicated that we might let the magnitude of the field strength be represented graphically by the number of lines of force drawn **normally** through unit area. It will be convenient to restrict such a representation to the case where the charge is in free space, or air. On this basis the flux density (lines per unit area) **in some medium other than air** would be given by the expression

$$N = K\mathcal{E}, \tag{5}$$

where N represents the number of lines drawn normally through each square centimeter. N is frequently referred to as the **electrostatic induction** or **electrostatic flux.** The lines representing $\mathcal{E}$ stand for **lines of force.** Lines representing N are known as **lines of induction.** In either case, the number of lines passing normally through unit area is indicated. **N indicates flux density**; this quantity is sometimes designated by the letter D. But by adopting the above indicated conventions, it is possible to have the number of lines remain the same if and when they chance to pass through several mediums in succession. Lines of flux are therefore **continuous,** even though they cross a bounding surface, and we **shall encounter such cases. Later we shall have occasion to speak of the total flux** over some given area. This total flux may be conveniently

expressed by either one of two equivalent relations, thus

$$\psi = NA = DA, \tag{6}$$

where ψ stands for the total flux, N the flux density, and A the area involved. D is synonymous with N.

The existence of electrostatic flux involves electronic displacement. In the case where a material medium, such as an insulator, separates two charges, it is found that, while there are few, if any, free electrons present, there is a slight displacement of the orbital electrons of the atoms constituting the dielectric. In fact, dielectrics placed in an electrostatic field do show surface charges. It is thus evident that some sort of transient electronic movement takes place. The situation might be graphically represented as shown in Fig. 3. In (a) the plates A and B are not

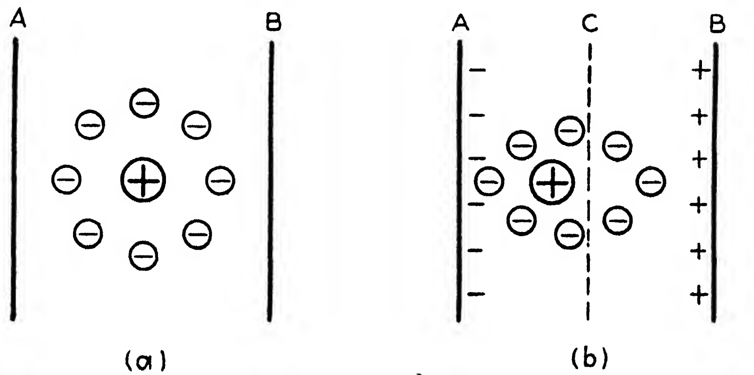


FIG. 3.—Electronic displacement due to an electrostatic field.

charged. In (b) the plates are charged as shown, with the result that at least some of the electrons of the dielectric undergo a change of position; there is, in short, a momentary electronic displacement across any plane such as C. If the charges are removed from A and B, the atoms of the dielectric will revert to their original electronic status. The physical extent of any displacement will be a function of the magnitude of the charges on the plates. It is this electronic **displacement** with which we are concerned. A dielectric thus acted upon by the forces which obtain in an electrostatic field is said to be **polarized**; and an atom thus distorted is called an **electric dipole** or an **electric doublet.** As the result of this polarization process, the field **within** the dielectric is somewhat diminished in magnitude. The extent of the decrease of the internal field strength depends upon the nature of the medium; and this physical property of a dielectric is what we designate by the constant K. **The dielectric constant is, accordingly, an index of the extent to which an insulating medium**

may be polarized. As we shall see later, the field within a **conductor** is zero. Therefore the dielectric constant, in such a case, would be infinitely great.

If one considers the case where the space separating two charges is **not** occupied by a material medium, the region may be thought of as being occupied with a universal medium (call it the ether, if you please) that possesses the property of elasticity, and will therefore be subject to a strain (distortion). **This strain constitutes a polarization of the ether.** The ether is present in the region between the charges even though a material medium is also present. The total displacement through unit area (the sum of the polarization of the ether and that of the material medium) is a vector quantity, and may therefore be represented by lines of induction or lines of flux. The magnitude of this electric induction normal to unit area is given by the factor N in Eq. (5).

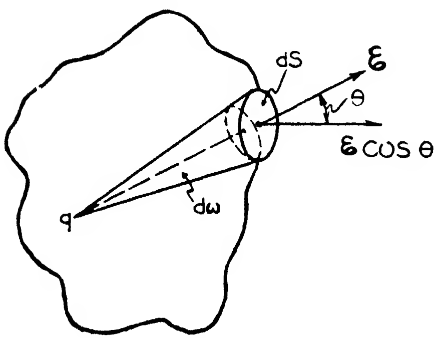


FIG. 4.—Gauss's theorem.

9. Gauss's Theorem. Having arrived at an understanding of what is meant by electric flux, we may now proceed to find a quantitative relation between the total normal induction (the flux which passes outward at right angles to the surface) over a closed surface and the magnitude of the charge which gives rise to it.

Assume any closed surface surrounding a charge q, as sketched in Fig. 4. We shall further assume that the space inclosed by the surface contains a dielectric whose constant is K. Let ds be any small area on this surface and $\mathcal{E}$ the electric intensity at a point on this elemental area, due to the charge q. We are interested in the **normal component** of this field intensity; this, obviously, would be $\mathcal{E} \cos \theta$. By Eq. (5) the normal flux (induction) through unit area would be given by

$$N = K(\mathcal{E} \cos \theta).$$

The normal flux through the area ds would then be

$$N_{ds} = K(\mathcal{E} \cos \theta)\, ds.$$

By Eq. (3) the above becomes

$$N_{ds} = K \frac{q}{Kr^2} (\cos \theta)\, ds$$

where r is the distance from the charge q to the surface area ds. This

reduces to

$$N_{ds} = q\,\frac{ds \cos \theta}{r^2}.$$

Now the expression $(ds \cos \theta)/r^2$ is equivalent to the solid angle $d\omega$ subtended at the point where the charge q is located by the area ds. Hence we may write

$$N_{ds} = q\left(\frac{ds \cos \theta}{r^2}\right) = q\,d\omega.$$

The **total** normal flux (displacement) over the **entire** surface would then be given by the surface integral of $q\,d\omega$, or

$$\psi_s = \int_S q\,d\omega$$

or

$$\psi_s = 4\pi q \qquad \text{(lines, normal, outward).} \tag{7}$$

This is a mathematical expression for Gauss's theorem which, in words, is to the effect that **the total normal flux (induction) across any surface inclosing a charge is numerically equal to 4π times the magnitude of the charge involved.** If, then, q be unity, it follows that 4π lines of electrostatic flux terminate on unit charge. This is a point to be kept in mind. If the surface incloses a number of charges q may be taken as the algebraic sum of the charges inclosed by the surface. It is customary to say that if q is positive the flux is directed outward, and if negative it will be directed inward. Obviously, if there be no resultant charge within the surface the total normal displacement (flux) will be zero. It may be shown that any charge outside the surface would not contribute to the total flux over the surface. The relation embodied in Eq. (7) will be found useful in connection with the solution of a number of problems which we shall encounter in our study of the laws of electrostatics.

10. Surface Distribution. Before dealing with several cases to which Gauss's theorem applies, it will be well to consider what is meant by the term "surface distribution." There is abundant experimental evidence, as well as analytical reasons, for a statement to the effect that, in the case of a charged conductor, the charges which are in electrical equilibrium are to be found only **on the outer surface.** Faraday, Webb, and others strongly electrified a relatively large metal inclosure but were unable to detect any evidence which would indicate the existence of charge within the inclosed region. The distribution of the charge on the surface of a conductor depends upon the geometrical form of the conductor, and also upon the presence and distribution of any surrounding

charges. This follows from the fact that all charges give rise to fields of electrostatic force and thus interact upon one another in conformity with the relation expressed in Eq. (2). Since we are dealing with **charges at rest,** the lines of force must at all points be normal to the conducting surface. If this were not so the field intensity would have a component tangential to the surface and **motion** of the charge would therefore result.

Quantitatively, what we may call the **average surface density** of a charge is given by the expression

$$\sigma = \frac{Q}{A}; \tag{8}$$

where Q is the total charge and A the area of the conducting surface. The surface density σ will be expressed in statcoulombs per square centimeter. The density **at any point** would be given by dQ/dA. In general, the surface distribution will not be uniform. In a few special cases, however, σ may approximate a constant value; such, for example, as in the case of a sphere far removed from other charges.

It is a matter of common experience that the surface density on a conductor will be greatest on those parts of the conductor which have the greatest surface curvature, approaching infinity in the case of sharp points.

It is difficult to account for this distribution analytically. Coulomb[1] experimentally investigated the elementary laws governing the distribution of electricity on conductors, but he found it extremely difficult to treat the general problem analytically. In fact, a complete analytical treatment has been attained in only a few special cases, one being that of the ellipsoid.[2] It is beyond the scope of this volume to review the mathematical analysis of this question, but it may be said that **the distribution of the charge is such that the resultant electrostatic field will be zero at all points within the conductor.** Later we shall see that the high surface density which obtains in the case of pointed conducting areas is of considerable practical significance.

11. Electric Intensity at Any Point Outside a Uniformly Charged Surface. In connection with the study of various instruments and

[1] Coulomb's fifth memoir, "Histoire d l'Académie," 1787.

[2] First and second memoirs of Poisson, "Memoirs de l'Institut," 1811. Green's Essay on the "Application of Mathematical Analysis to the Theories of Electricity and Magnetism," 1828. "Papers on Electrostatics and Magnetism," by Sir William Thomson (Lord Kelvin), Vol. XV, p. 178. See also the second paper in the same collection. The student will find the consultation of original sources a fascinating and profitable pursuit.

devices, it will be found useful to have available certain analytical tools which involve the concept of field intensity. With this end in view, we shall next derive expressions giving the value of the electric intensity, or field strength, in certain special cases.

Let us first develop an expression for field intensity at any point **outside** a uniformly charged sphere. To do this let us assume a sphere as shown in cross section in Fig. 5. Our problem is to find the field strength at a point p. It is possible to derive two expressions for Ψ_S, the total normal flux; we may then equate these and solve for the field intensity.

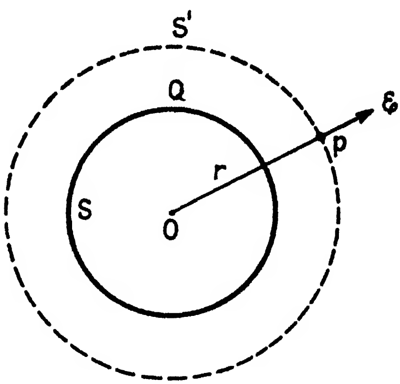


FIG. 5.—Field outside of a charged sphere.

Let p be any point distant op from the center of a conducting spherical surface S having a charge Q. For convenience we will assume that the sphere is electrically isolated, thus making the flux normal to the surface. Draw a spherical surface S' concentric with that of the sphere and passing through the point p. The flux which leaves the charged sphere will pass through S'.

Now the total normal flux over this hypothetical sphere will be given by the expression

$$\psi_{S'} = K \times \mathcal{E} \times \text{area of } S'$$
$$= 4\pi\overline{op}^2K\mathcal{E}.$$

By Gauss's theorem

$$\psi_{S'} = 4\pi Q.$$

Equating, we have

$$4\pi Q = 4\pi\overline{op}^2K\mathcal{E}$$

$$\mathcal{E} = \frac{Q}{K\overline{op}^2} \qquad \text{dynes/statcoulomb.} \tag{9}$$

Infinitely near the surface, op becomes practically identical with the radius of the sphere. But

$$Q = \text{area} \times \text{surface density} = 4\pi r^2\sigma,$$

where r is the radius of the sphere S. Hence we may write

$$\mathcal{E} = \frac{4\pi\sigma}{K} \qquad \text{dynes/statcoulomb} \tag{10}$$

which gives the field strength **infinitely near** a charged sphere.

From Eq. (9) it is evident that for any point outside a uniformly charged sphere the field intensity is the same as if the entire charge on the conducting sphere were concentrated **at its center.**

Problem. An insulated conducting sphere 10 cm in diameter carries a charge of 100 statcoulombs uniformly distributed over its surface. (*a*) What is the surface density of the charge? (*b*) What is the field strength 10 cm in air from the surface of the sphere? (*c*) What is the field intensity infinitely near the surface?

Solution. (*a*) The area of the sphere would be $4\pi \times 5^2$, or 100π cm². By Eq. (8)

$$\sigma = \frac{100}{100\pi} = \frac{1}{\pi} = 0.318 \text{ statcoulombs/cm}^2.$$

(*b*) By Eq. (9)

$$\mathcal{E} = \frac{100}{15^2} = 0.444 \text{ dynes/unit charge.}$$

(*c*) In this case r approaches the radius of the sphere as a limit; hence

$$\mathcal{E} = {}^{100}\!/_{25} = 4 \text{ dynes/statcoulomb.}$$

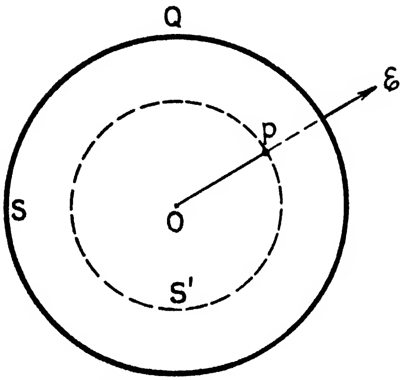


FIG. 6.—Field within a charged sphere.

As a check we might utilize Eq. (10) and get

$$\mathcal{E} = (4\pi)\frac{1}{\pi} = 4 \text{ dynes/statcoulomb.}$$

12. Field Intensity at Any Point Inside a Uniformly Charged Shell. In this instance we adopt the same procedure as in the previous case. In Fig. 6, the point under consideration is at p **inside** the hollow spherical conductor S. The sphere is again assumed to be electrically isolated, thus making the surface distribution uniform. The hypothetical sphere S' is concentric with S and includes p on its surface. As before, the total normal flux over S' would be

$$\psi_{S'} = K\mathcal{E} \times \text{area of } S' = 4\pi\overline{op}^2K\mathcal{E}.$$

Again applying Gauss's theorem, the flux through S' would be

$$\psi_{S'} = 4\pi Q.$$

Equating, we have

$$4\pi\overline{op}^2K\mathcal{E} = 4\pi Q,$$

which yields

$$\mathcal{E} = \frac{Q}{K\overline{op}^2}.$$

But, since there is no charge **within** the sphere S', the value of Q is zero. It accordingly follows that the field intensity $\mathcal{E}$ will also be zero. In words, this means that **there is no electric field inside a uniformly charged spherical shell.** Actually, as pointed out in Sec. 9, there is no field within a charged conducting shell, whatever its shape.

13. Electric Intensity at any Point Outside a Uniformly Charged Cylinder. Let us assume that we have an infinitely long cylinder having a uniform charge of q units per centimeter of length. The problem is to find an expression for the field strength at any point p outside the con-

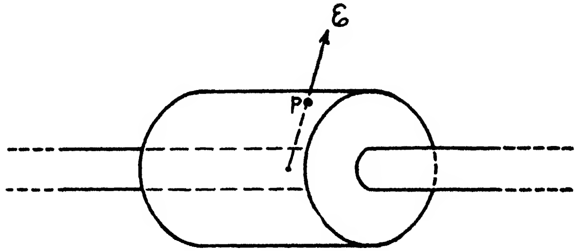


FIG. 7.—Field outside of a charged cylinder.

ductor. As shown in Fig. 7, a coaxial cylindrical surface C of unit length is drawn so that its surface includes the point p. The planes forming the ends of the construction cylinder C are perpendicular to the common axis. The radius of cylinder C will be taken to have a value r. The flux will be radial with respect to the conductor and to the surrounding cylinder C, and hence parallel to the planes at the ends of C. There will, therefore, be no normal flux through these planes. We have only to consider, then, the area of the curved surface. The area of this surface will be $2\pi r$. The total normal flux through the cylinder C will be given by

$$\psi_C = K\mathcal{E} \times \text{area of } C = 2\pi r K\mathcal{E}.$$

The application of Gauss's theorem to the problem gives us

$$\psi_C = 4\pi q.$$

Equating these two values for the total normal flux over C, we get

$$2\pi r K\mathcal{E} = 4\pi q.$$

Therefore

$$\mathcal{E} = \frac{2q}{Kr} \qquad \text{dynes/statcoulomb,} \tag{11}$$

where **q is the charge per unit length** and r the distance from the point in question to the axis of the charged cylindrical conductor. It will be

noted that the radius of the charged cylinder itself does not enter into the case. Further, Eq. (11) shows that, **for points outside, the charge acts as if it were concentrated on the axis of the cylinder.** In the case of a conductor of finite length, Eq. (11) gives the magnitude of the field strength to a first approximation if the point is at a distance from the conductor that is small compared to the length of the conductor.

In a similar manner, it may be shown that **the electric intensity is zero at any point inside a uniformly charged cylindrical shell.** It is found that even though the conducting surface is irregular in shape there is no electrostatic field on the inside. In practice it is frequently necessary to shield conductors and instruments from the fields due to extraneous charges. By placing the conductors or apparatus in a metal inclosure they may be shielded from the effects of outside fields. It is not, in all cases, essential that the shielding surface be continuous; frequently a wire screen will answer the purpose. For reasons that we shall see later, it is usually necessary to connect the shield to the earth.

Problem. What is the field intensity in air at a point 3 cm from a cylindrical conductor 10 m in length which carries a charge of 5×10^{-5} coulombs/cm of length?

Solution. Bearing in mind that the charge in coulombs must be expressed in statcoulombs, and that the charge per unit length must be used, we may write

$$\mathcal{E} = \frac{2 \times 5 \times 10^{-5} \times 3 \times 10^{9}}{3 \times 10^{3}} = 100 \text{ dynes/statcoulomb.}$$

14. Electric Intensity at any Point near a Uniformly Charged Infinite Plane. In Fig. 8 let BD represent a plane of infinite extent. Assume that this plane is charged uniformly **on both sides,** say positively, and that

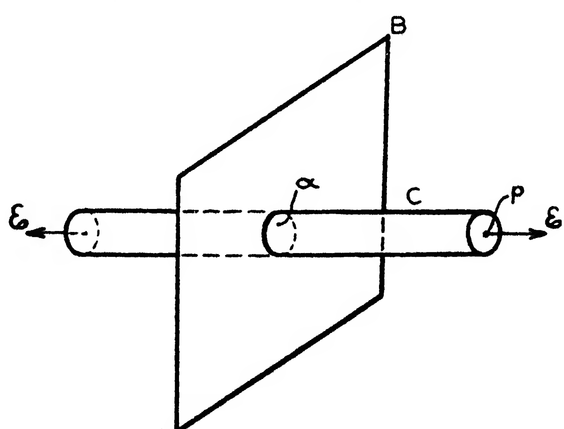


FIG. 8.—Field due to a charged plane surface.

the surface density is everywhere equal to σ. Draw a hypothetical right cylinder C cutting the plane as shown. Let the area of a be 1 cm². At a point not too far from the plane, the flux lines will be parallel and at right angles to the plane's surfaces, and outwardly directed from both sides of a. Then, since the flux is parallel to the axis of the cylinder, only the ends of the cylinder will be cut by the flux. On the basis of the above indicated assumptions with regard to the significance of σ, the total charge inclosed by the Gaussian surface will be 2σ. By Gauss's theorem, then, the total normal flux will be

$$\psi_C = 4\pi \times 2\sigma = 8\pi\sigma.$$

The total flux would also be given by

$$\psi_C = K\mathcal{E} \times \text{area}.$$

Equating we have

$$K\mathcal{E} \times \text{area} = 8\pi\sigma.$$

Since flux is emerging from both ends of the cylinder, the total area involved would be 2 cm². Accordingly, the last equation reduces to

$$\mathcal{E} = \frac{4\pi\sigma}{K} \qquad \text{dynes/statcoulomb.} \tag{12}$$

Equation (12) embodies a relation which is frequently referred to as **Coulomb's law.** In words, this equation means that, in air, the field intensity at any point **very near** a charged surface is numerically equal to the product of 4π and the charge on unit area **on one side** of the conductor. It is to be noted that, on the assumption that the flux is normal to the conducting surface, the magnitude of the field intensity is **independent of the distance of the point from the plane.** It is interesting to observe that this is exactly the same expression that was derived for the case of the charged sphere: Compare Eqs. (10) and (12). In fact, this is a general relation that we shall presently find useful.

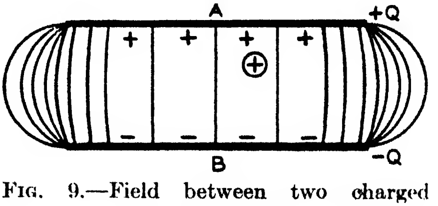


FIG. 9.—Field between two charged parallel plates.

15. Force Exerted by a Charged Plate on a Second Parallel and Oppositely Charged Plate. In practice, one sometimes encounters a situation where it becomes necessary to be able to compute the magnitude of the electromechanical force exerted by a charged conducting plate on a second parallel and oppositely charged plate. The case is diagrammatically shown in Fig. 9. If a charge of $+Q$ statcoulombs is placed on plate

A and an equal negative charge on B, the charges will reside entirely on the adjacent sides, as shown. The surface density, σ, in each case will be Q/area. Except near the edges, the lines of flux will be normal to the plates and hence parallel to one another. These lines may be thought of as originating on the positive charges and terminating on the corresponding negative charges. It has been shown (Sec. 8) that 4π lines leave each unit charge; therefore there will be $4\pi\sigma$ lines per square centimeter between the plates. The field intensity $\mathcal{E}$, at any point between the plates, therefore, will be $4\pi\sigma$ dynes per unit charge. It is to be noted that the presence of the charges of opposite sign **does not increase the existing flux.** A unit test charge (positive) placed anywhere between the plates in any dielectric medium will, therefore, be acted upon by a force of $4\pi\sigma/K$ dynes. **Half of this force, or $2\pi\sigma$ dynes, will be due to the repulsion of the positive charge on A and the other half to the attraction due to the corresponding negative charge on B.** Since, then, each unit area of one of the charged plates, say A, exerts a force of $2\pi\sigma$ dynes on any unit charge, and since a charge of σ units resides on each unit area of the other plate B, the force experienced by that quantity of charge will be

$$F = \frac{(2\pi\sigma)\sigma}{K} = \frac{2\pi\sigma^2}{K}. \tag{i}$$

We have already seen (Sec. 14) that the field intensity in the region of a charged surface is given by the expression

$$\mathcal{E} = \frac{4\pi\sigma}{K}. \tag{ii}$$

Solving for σ in this equation, we obtain

$$\sigma = \frac{\mathcal{E}K}{4\pi}. \tag{iii}$$

Substituting this value of σ in (i), we have as a working relation for the force experienced by either of the charged plates

$$F = \frac{\mathcal{E}^2K}{8\pi} \qquad \text{dynes/unit area.} \tag{13}$$

This is a general and an important relation; it will give the force on unit area of any charged flat conductor. The total force acting on the plate would be given by the product of the above expression and the area of one of the plates.

16. Energy per Unit Volume of the Medium. It has already been pointed out that the medium surrounding a charge is in a state of strain. This holds even when no material medium occupies the field. In order to produce a state of strain, work must be done; energy must be expended.

The medium surrounding a charge must, therefore, be the seat of potential energy. It will be found useful to be able to evaluate this energy.

From Eq. (13) we see that the force (electrostatic) acting on unit area of a charged surface is $K\mathcal{E}^2/8\pi$ dynes. If, then, we move unit area of such a charged surface a distance dx, the work done would be

$$dw = \left(\frac{K\mathcal{E}^2}{8\pi}\right) dx.$$

If the charged surface were moved unit distance, thus increasing the volume by unity, the total work done would be

$$\int dw = \frac{K\mathcal{E}^2}{8\pi} \int_{x=0}^{x=1} dx$$

or

$$W = \text{P.E.} = \frac{K\mathcal{E}^2}{8\pi} \qquad \text{ergs/cm}^3. \tag{14}$$

This, then, should give the energy content of 1 cm^3 of the medium.

Problem. What is the energy per unit volume at a point 0.5 cm from a plate whose dimensions are 10×10 cm, the charge on the plate being 500 statcoulombs? The surrounding medium is glass having a constant of 5.

Solution.

$$\sigma = \frac{Q}{A} = \frac{500}{10 \times 10} = 5 \text{ statcoulombs/cm}^2.$$

$$\mathcal{E} = \frac{4\pi\sigma}{K} = \frac{4\pi \times 5}{5} = 4\pi \text{ dynes/unit charge.}$$

$$\text{P. E.} = \frac{K\mathcal{E}^2}{8\pi} = \frac{5 \times (4\pi)^2}{8\pi} = 10\pi \text{ ergs/cm}^3.$$

17. Refraction of Lines of Flux. At present we are studying charges at rest (electrostatics); later we shall consider the effect of electrons in motion. When charges are at rest the concomitant field has a fixed value at any given point. If the charges move, the accompanying field will vary in value; and if the motion is periodic, there will be developed what we shall later call electromagnetic waves. One aspect of such a wave disturbance consists of periodic variations in the value of the field intensity $\mathcal{E}$. These harmonic disturbances move and in so doing may encounter strata of mediums having various dielectric constants. Interesting and important phenomena often appear at the **boundary between two such mediums.** In view of this fact, we may with profit consider briefly what happens when electrostatic flux extends across a boundary between two dielectric mediums having different constants.

It may be shown that at the boundary between two different dielectric mediums in an electrostatic field (Fig. 10) two conditions must obtain:

(1) the **tangential** component of the electric intensity $\mathcal{E}$ in the first medium must equal the **tangential** component of the intensity in the second medium; and (2) the **normal** component of the flux N in the first medium must equal the corresponding component in the second medium. The first condition cited above may be expressed thus:

$$\mathcal{E}_1 \sin \theta_1 = \mathcal{E}_2 \sin \theta_2, \tag{i}$$

and the second condition as

$$K_1\mathcal{E}_1 \cos \theta_1 = K_2\mathcal{E}_2 \cos \theta_2. \tag{ii}$$

From these two boundary conditions it is possible to find the **change in direction of the field** as one passes from the first medium to the second. By dividing (i) by (ii), and remembering that $\mathcal{E}_1 = \mathcal{E}_2$ (Sec. 7), we get

$$\frac{K_1}{K_2} = \frac{\tan \theta_1}{\tan \theta_2}. \tag{15}$$

Equation (15) thus becomes an analytical tool by means of which one

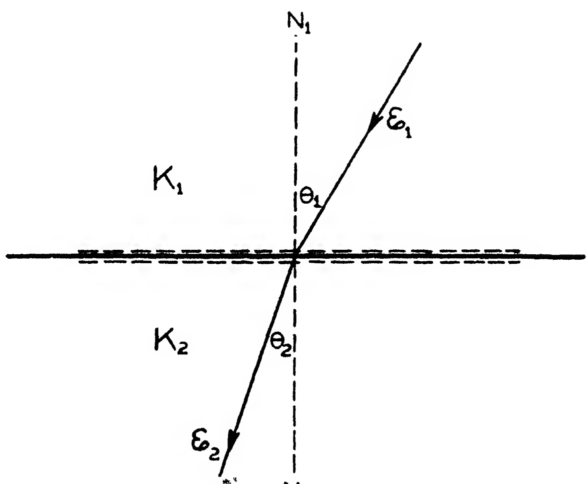


FIG. 10.—Refraction of electrostatic flux.

may evaluate the refraction that electrostatic flux undergoes in extending across a boundary between two dielectric mediums. The distortion of the field can therefore be determined. From Eq. (15) it is evident that if $K_1 > K_2$, $\theta_1 > \theta_2$; that is, when flux passes from a medium of one dielectric constant to another having a smaller value it is bent toward the normal, and vice versa.

PROBLEMS

1. How many electrons are present in a charge of 100 statcoulombs?

2. What is the total flux emanating from a concentrated charge of 2×10^{-5} coulomb?

3. Two concentrated unlike charges of 100 statcoulombs each are separated by a distance of 10 cm. What is the magnitude of the force, in grams, tending to draw them together? What would be the value of the force if the charges were immersed in oil having a dielectric constant of 2.2?

4. An atom of hydrogen consists of a proton and a single electron. If the average distance between these two entities is taken to be 10^{-8} cm, and if we assume that the inverse square law holds (which may not be the case), calculate the magnitude of the force of attraction between the proton and the electron.

5. What will be the field intensity at a point 25 cm from a concentrated charge of 3×10^{-6} coulomb? What electromechanical force will be experienced by a charge of 2×10^{-5} coulomb at the point designated?

6. Three point charges are located on the circumference of a circle at equal distances apart, the diameter of the circle being 10 cm. The magnitude and sense of the charges are $+6$, $+10$, and -12 statcoulombs, respectively. What is the field intensity at the center of the circle?

7. Two small spheres, each having a mass of m grams, are suspended by silk threads from a common point. The supporting threads are l centimeters long; the spheres take up a position d cm apart. Develop an expression which will give the magnitude of the charge on one of the spheres as a function of l, d, and g (acceleration due to gravity). Assume the spheres to have equal like charges.

8. Utilizing the relation developed in Prob. 7, find the charge on each sphere when two such bodies stand 10 cm apart. The threads are 60 cm long, and each body weighs 0.6 gm. Take g to be 980 cm/sec².

9. What is the field intensity at a point 50 cm from a sphere that has a charge of 5×10^{-5} coulomb? Will it make any difference whether the charge on the sphere is positive or negative?

10. What will be the magnitude of the force, in grams, acting on a body carrying a charge of 500 statcoulombs if this charged body is located 0.5 cm from a cylindrical conductor which is 2 m in length? The conductor bears a charge of 2×10^{-6} coulomb, and the surrounding medium is air.

11. What total flux will issue from a conducting plate measuring 10×10 cm which bears a charge of 5×10^{-5} coulomb, the surrounding medium being air? What would the flux be if the medium were carbon dioxide gas at a pressure of 20 atm? The field distortion at the edges may be neglected.

12. In the previous problem, a second conducting plate of the same size is placed near and parallel to the first plate. This second plate is given an equal but opposite charge. What will be the force of attraction between the plates, in the two cases?

13. Under the conditions cited in the last problem above, what will be the energy content of the intervening medium when the plates are 2 mm apart? Will the energy content increase or decrease if the gas pressure is increased?

14. Again referring to Prob. 12 above, how much work would be done in moving the plates 1 cm farther apart?

15. Suppose the direction of a certain electrostatic flux in rubber makes an angle of 45° with the interface as it passes from rubber into glass. What will be the angle of the flux direction in the second medium? Take K for the rubber to be 2.5 and for the glass to be 6.

16. Assume that we have an electron moving in free space in a straight line with a velocity of 3×10^8 cm/sec. At a certain point in its path it encounters an electrostatic field due to a charged plate whose surface density is 100 statcoulombs/cm^2. If the direction of the field is at right angles to the original line of motion the electronic path will tend to be circular. Under the circumstances indicated, what will be the radius of the circle?

17. What will be the magnitude of the mechanical force, in grams, acting on a body which carries a charge of 2×10^{-6} coulomb when located 10 cm from a cylinder which bears a charge of 5×10^{-5} coulomb per unit length?

18. Three positive charges are located at points equally distant apart on the surface of a hypothetical sphere. What will be the field strength at the center of the sphere if the charges are 100 statcoulombs each, and the radius of the sphere is 10 cm? What will be the field intensity at a point 50 cm from the center of the sphere?

19. Point charges of +25, −50, +75, and −100 statcoulombs, respectively, are located at the corners of a 10×10 cm rectangle. What is the field strength at the center of the figure?

20. Under the conditions specified in the last problem, locate a point where the field is zero.

21. Two long insulated, conducting, concentric cylinders have radii of 5 and 6 cm, respectively. The inner cylinder is given a positive charge of 100 statcoulombs and the outer one an equal negative charge. Find the field strength (*a*) at a point on the common axis, (*b*) at a point 5.5 cm from the axis, and (*c*) at a point 10 cm from the axis.

CHAPTER III

MAGNETOSTATICS

18. Historical. Certain aspects of the phenomenon which we broadly designate by the term magnetism were known to the ancients. The use of iron as a metal reaches back to a very early date; and the first knowledge concerning magnetism probably had its origin in connection with the working of the iron ore which we know as magnetite[1] (Fe_3O_4). This particular ore, so tradition has it, was found in or near Magnesia, in Asia Minor, and Lucretius suggested that the term "magnet" was derived from that geographical name. Other writers ascribe the origin of the term to other sources. Several of the ancient writers, including Lucretius, Pliny, and Socrates, mention some of the more obvious facts in connection with the behavior of the magnetic ore; but magnetic polarity and the concomitant phenomenon of attraction and repulsion between poles apparently were unknown to Greek antiquity. Natural magnets, consisting of pieces of magnetite, were originally called by various names, among which was the designation "loadstone" (sometimes spelled lodestone). The loadstone itself appears to have been utilized as a crude compass before the beginning of the Christan era, but it was probably not until the Middle Ages that it became known that permanent or "artificial" magnets could be produced by the simple process of bringing pieces of steel into proximity with natural magnets. In 1269 the writing of Peter de Maricourt indicated that he was acquainted with magnetic polarity, the law of attraction, and with other important properties of magnets. Indeed, it is to this investigator that we owe the term "magnetic poles."

Magnetism as a definite branch of science had its beginning with the appearance in 1600 of an epoch-making treatise entitled "De magnete" by Dr. Gilbert, an eminent English physician. In this classical volume Dr. Gilbert reviewed the then known facts concerning magnetism and in addition presented many original and important observations of his own. It is to him that we owe the suggestion that the earth acts as a great natural magnet. Indeed, the terminology associated with the subject of magnetism is due largely to Gilbert.

Though we now know that the property we call magnetism can be imparted either temporarily or permanently to a considerable number of

[1] Magnetite is widely distributed, being found in the Ural Mountains, Scandinavia, Finland, Canada, and in the states of New Jersey, Pennsylvania, and New York.

substances; and though we have made extensive use of magnetic forces; we as yet know very little about the true nature of this strange but highly important phenomenon. Various explanations have been offered to account for the existence of magnetism, but as yet no single theory has been found to be entirely satisfactory. There is evidence which tends to indicate that magnetism has its genesis in the atoms or the molecules of certain substances. But why in some atoms or molecules and not in all? In more recent times the question is being approached from the standpoint of the electron theory of matter. Possibly the motions of the orbital electrons may in some way account for magnetic properties. We shall return to this question after we have considered the properties of moving electrons (the electric current). For the present we may address ourselves to a study of the known laws that govern the interaction between centers of magnetism.

19. Magnetic Poles and the Laws of Force. Magnetostatics has to do with the study of the laws governing the force interactions of magnetic poles at rest. While there are many striking similarities between the behavior of magnetic poles and electric charges, there are also a number of important differences. These similarities and differences will become apparent as we proceed.

We have just used the term **magnetic poles.** What does that expression signify? A piece of magnetite, or a piece of steel to which the property of magnetism has been imparted, will be found to have **at least** two centers where the magnetism is most pronounced. The regions are not, in general, at the geometrical ends of the piece of magnetic material; indeed their exact location is difficult, if not impossible, to determine. These regions, or centers, of maximum magnetic intensity are called the poles of the magnet. A line joining the poles is referred to as the magnetic axis. If a magnet is suspended it tends to set itself so that its magnetic axis is roughly parallel to the north-south direction. One is called the north-seeking and the other the south-seeking pole. In electrostatics, as we have seen, it is possible to have an isolated charge, **so far as a given body is concerned**; but, in dealing with magnetism, at least two poles of opposite character are always present **in the magnetic body itself.** There may, however, be many poles present in a body, such as in the case of a long steel tape or wire, as we shall see later. While we cannot isolate a magnetic pole, for purposes of analytical convenience, we do speak of a single pole. If the distance between the two poles of an ordinary magnet is relatively great, either one of the poles may be thought of, for some purposes, as isolated.

As in the corresponding case in electrostatics, there are two laws of force action involving magnetic poles. The **first law** is to the effect that

like poles repel and unlike poles attract. We mean by "like poles" those that tend to point in a like direction if the magnet is free to rotate about a vertical axis.

The second basic relation may be expressed symbolically thus,

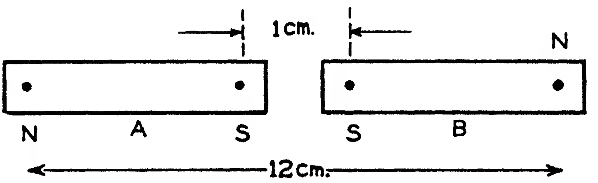
$$F = \frac{mm'}{\mu d^2}, \tag{16}$$

where F is the force of attraction or repulsion manifest between the two poles involved, m and m' the respective pole strengths, d the distance separating the poles in question, and μ a factor which depends for its value on the nature of the intervening medium. This factor, called permeability, is **not** a constant, as is K in the corresponding relation in electrostatics. We shall find that the permeability of a given material is determined to some extent by a certain condition which will be touched upon shortly. Coulomb investigated the relation embodied in Eq. (16) as well as the corresponding relation in electrostatics [Eq. (2)]. In doing this he studied the force action between the poles of two long and very slender magnets. Since a magnetic pole occupies a rather indefinite volume, high accuracy of measurement could not be attained. But he found the inverse square law to hold, at least approximately. Later Gauss, by an indirect method, gave a more rigorous proof of this relation. Equation (16) is frequently referred to as Coulomb's law of magnetic poles. The similarity between the expressions embodied in Eqs. (2) and (16) is striking. Here, as in electrostatics, we encounter an inverse square law; here we have pole strength corresponding to quantity of charge, and the factor permeability replacing dielectric constant.

In considering the concept of pole strength we are dealing with a **quantity,** and hence it becomes necessary to agree upon a unit in which such a quantity may be expressed. As in the case of electrostatics, we may define that unit **in terms of mechanical force.** Referring to Eq. (16) it may be said that **unit magnetic pole is one whose pole strength is such that when placed at a distance of one centimeter in free space from a like pole of equal strength it will experience a repelling force of one dyne.** Thus pole strength will be expressed in terms of the unit pole just defined. The unit of magnetic pole strength has no special name; however, it serves as the basis of a system of electrical units referred to as the "electromagnetic system." It thus becomes evident that we have two fundamental systems of electrical units, one based on unit charge (the electrostatic system), and another based on unit pole. It will be pointed out, as we proceed, that there is a third system of electrical units (the engineering or practical system), each member of which bears a definite relation to the corresponding electrostatic and magnetic units.

The concept of pole strength can be made more real if we apply it to a definite case, such as the following.

Problem. Suppose that we have two bar magnets positioned as shown in the accompanying sketch. If we assume the pole strength of magnet A to be 200



units and the pole strength of B to be 250 units, what will be the magnitude of the force tending to move the magnets?

Solution. The force of repulsion between the S poles will be

$$F_S = \frac{200 \times 250}{1^2} = 50{,}000 \text{ dynes.}$$

The force of repulsion between the N poles will be

$$F_N = \frac{200 \times 250}{12^2} = 347 \text{ dynes.}$$

The force of attraction between each pair of N-S poles will be

$$F_{NS} = \frac{200 \times 250}{(6.5)^2} = 1{,}181 \text{ dynes.}$$

The resultant force will be one of repulsion as given by

$$50{,}000 + 347 - 2 \times 1{,}181 = 47{,}985 \text{ dynes.}$$

20. Magnetic Field and Field Intensity. The interpole force action between magnetic poles indicates that the region surrounding a pole possesses certain characteristics that distinguish it from ordinary free space, as was the case in dealing with electric charges. **Any region in which a magnetic effect may be detected is referred to as a magnetic field.** The distance at which such effects may be detected is great, but, because of the operation of Coulomb's inverse square law, the practical extent of the field is quite limited.

Reverting to Eq. (16) again, if we make one of the poles, say m', have a magnitude of unity, then the force which it would experience when in the field of a pole whose value is m and located at a distance d would be given by the expression

$$F = \frac{m}{\mu d^2}.$$

This force per unit pole is known as **magnetic field strength** and is usually designated by the letter H. Accordingly, the above equation is written

$$H = \frac{m}{\mu d^2}. \tag{17}$$

In air the permeability factor becomes unity. Magnetic field strength is expressed in dynes per unit pole, or **oersteds,** in honor of Hans Christian Oersted, whose name we shall encounter in another important connection. While it is convenient to express field intensity in terms of a mechanical force, it should be clearly understood that in reality we are referring to a **property of the field** that gives rise to a mechanical force if a magnet is located at the point involved. Because it is expressed in terms of force, it is a vector quantity, as was the case in connection with electrostatic field strength.

If a given pole is located in a field of H oersteds, it will be subjected to a force given by the expression

$$F = mH \qquad \text{dynes.} \tag{18}$$

Problem. What is the magnetic field intensity in air at a point 10 cm from a pole of 500 units? To what mechanical force will a pole of 200 units be subjected if placed at that point?

Solution. Applying Eq. (17) we find that the field intensity is

$$H = \frac{500}{10^2} = 5 \text{ oersteds.}$$

The magnitude of the mechanical force, under the conditions cited, may be determined by the use of Eq. (18), thus

$$F = 200 \times 5 = 1{,}000 \text{ dynes.}$$

21. Field Intensity at a Point Near a Magnet. Thus far in our discussion of the magnetic field we have confined our attention to the effect of a single pole. Let us now examine what the situation is when the effect of both poles of a magnet is taken into consideration.

Let us assume that we have a bar magnet and that we wish to determine the field intensity at any point in its field. In Fig. 11, the resultant field H at p will be the vector sum of the fields due to the N-seeking and the S-seeking poles. These components are designated by H_N and H_S, respectively. The distance **between the pole centers** is represented by l. To simplify the discussion we will assume that $m = m'$, and that the medium is air.

With the aid of a well-known trigonometrical relation, we may write

$$H = \sqrt{H_N^2 + H_S^2 + 2H_NH_S \cos\theta} \tag{19}$$

and

$$\cos \theta = \frac{H^2 + H_S^2 - H_N^2}{2HH_S}.$$

By Eq. 17, we may write

$$H_N = \frac{m}{d_1^2}$$

and

$$H_S = \frac{m'}{d_2^2}.$$

From basic relations

$$\cos \theta = - \cos \alpha.$$

In order to be able to solve Eq. (19) and thus evaluate H, it will obviously be necessary to know the pole strength of the magnet, the distance l between the pole centers, and the magnitudes of d_1 and d_2. In any given

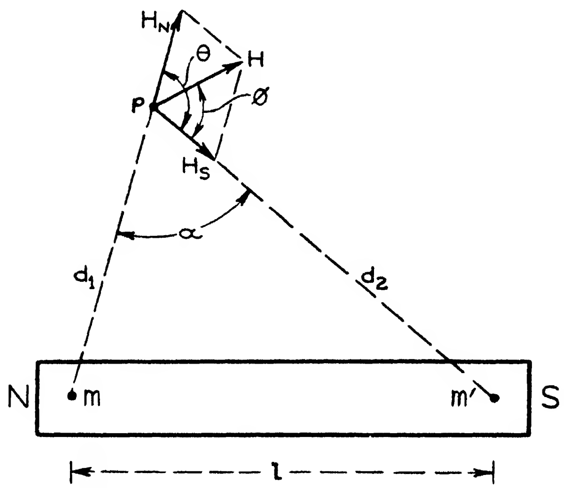


FIG. 11.—Field intensity at a point near a magnet.

case the magnitude of m can be determined experimentally, and the value of l can be arrived at indirectly by finding the product of m and l, as outlined in Sec. 23. Once these data are at hand, the magnitude and direction of the resultant field strength at any point in the region of the magnet can be computed. In the case of thin cylindrical bar magnets the value of l will be, roughly, five-sixths of the physical length of the magnet.

If one thinks of a magnet as being very short, and its poles as being concentrated at points exactly at the ends, we have what is referred to as a **magnetic dipole.** Such a concept is useful in connection with magnetic shells. The principles outlined in the above discussion are applicable to dipoles.

Problem. Referring to Fig. 11, if one assumes $d_1 = 5$ cm, $d_2 = 10$ cm, $l = 10$ cm, and $m = 100$ units, what will be the value of the field intensity at the point p? What is its direction?

Solution.

$$H_N = \frac{100}{5^2} = 4 \text{ oersteds}$$

$$H_S = \frac{100}{10^2} = 1 \text{ oersted}$$

$$\cos\theta = -\cos\alpha$$

$$\cos\alpha = \frac{5^2 + 10^2 - 10^2}{2 \times 5 \times 10} = 0.25$$

$$H^2 = 4^2 + 1^2 + 2 \times 4 \times 1(0.25)$$

$$H = 4.36 \text{ oersteds}$$

$$\cos\phi = \frac{(4.36)^2 + 1^2 - 4^2}{2 \times 4.36 \times 1} = 0.459;$$

hence

$$\phi = 63°, \text{ nearly.}$$

22. Lines of Force. Magnetic field intensity may be represented diagrammatically by lines between magnetic poles in a manner similar to the practice followed in representing the electric field. The lines are so drawn that at any given point, such as p in Fig. 12, they indicate the

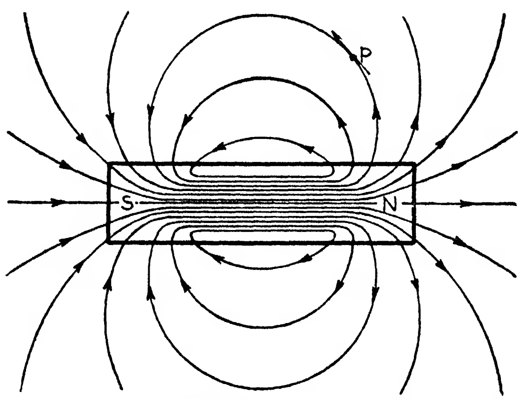


FIG. 12.—Lines of magnetic flux in and about a magnet.

direction of the force which would act on positive test pole if placed at that point. For convenience these lines are thought of as originating at the N-seeking pole and reentering the magnet at the S pole. Magnetic lines of flux, however, differ from electric lines in that they are considered to be **continuous,** threading through the body of the magnet between the S and N poles, as sketched in Fig. 12.

In addition to indicating the direction of field intensity, it is also customary to give to these lines an added significance. In the preceding section reference was made to the magnitude of the field strength. In

dealing with magnetic fields, it is convenient to adopt the convention that the **magnitude of the field intensity at a point shall be represented by the number of lines per unit area,** this area to include the point in question and be normal to the direction of the lines of force. If, for example, we have a magnetic field whose intensity is 10 oersteds, it would be understood that there are 10 lines per square centimeter.

By the same line of reasoning as followed in Sec. 9 it may be shown that there are $4\pi m$ lines of force emanating from a magnetic pole whose strength is m.

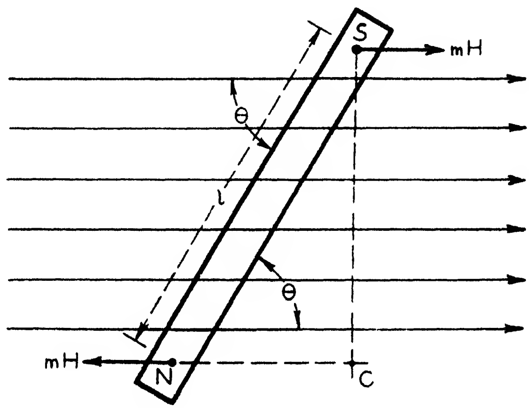


FIG. 13.—Forces acting on a bar magnet free to rotate about its center.

23. Magnetic Moment. It has already been pointed out that it is difficult to determine the exact points in a magnet which shall be taken as the "poles," or magnetic-force centers. Because of this uncertainty, the solution of a number of problems encountered in the study of magnetism is greatly facilitated by the use of a quantity known as **magnetic moment.** (At this point the student should review the subject of "moments" in general; it is important to have a clear understanding of the significance of that concept.) The moment of a magnet will be found to be a product of pole strength and the distance between the centers of magnetism. Let us see in what terms this quantity may be expressed.

If we suspend a magnet NS of pole strength m in a uniform magnetic field H, and in the position shown in Fig. 13, it will be acted upon by a couple tending to turn it into a position such that its major axis will be parallel to the direction of the field. The magnitude of the torque (moment of the couple) thus developed will be given by the expression

$$\begin{aligned} L &= mH(SC) \\ &= mHl \sin \theta, \end{aligned}$$

where l is the distance between the pole centers.

Now if an outside mechanical couple is applied to the system in such a manner that the magnet is held in a position perpendicular to the direction of the field θ becomes 90°, and hence $\sin \theta = 1$. Under these circumstances the applied torque would equal the magnetic torque and the last equation becomes

$$L' = mlH,$$

where L' is the magnitude of the applied torque. If the conditions are so arranged that the field intensity H is unity, this equation reduces to

$$L' = ml.$$

This product ml is the **magnetic moment.**

Bearing in mind the condition giving rise to the last equation, it is evident that, **numerically, the magnetic moment of any magnet is equal to the moment of the couple (torque) which is acting on a magnet when in a position such that its major axis is perpendicular to the direction of a uniform field whose intensity is unity.** The letter M is commonly used to designate the product ml and hence our final expression for magnetic moment is

$$M = ml. \tag{20}$$

Thus we see that, while l cannot be definitely determined, magnetic moment is a quantity which can be accurately measured and expressed in terms of a torque.

24. Intensity of Magnetization. Another concept which we shall find useful in dealing with magnetic problems is that of **intensity of magnetization.** Assuming that we have a body which is uniformly magnetized, **intensity of magnetization may be defined as the ratio of the magnetic moment to the volume.** This may be expressed as

$$I = \frac{M}{v} = \frac{ml}{al} = \frac{m}{a}, \tag{21}$$

where I is the intensity of magnetization and a the cross-sectional area. From the last equality in Eq. (21) it is evident that **intensity of magnetization may be defined as the pole strength per unit area.** In this respect, intensity of magnetization corresponds to electrostatic surface density and for that reason the letter σ is sometimes used instead of I to designate this quantity.

Actual magnets do not exhibit uniform magnetization throughout their entire volume. It is therefore customary to consider the intensity of magnetization **at a point,** and to express the quantity in the form

$$I = \frac{dM}{dV}. \tag{22}$$

25. Magnetic Shell. The quantity defined in the preceding section enters into the important concept which is referred to as a **magnetic shell.** If we imagine a thin sheet of material so magnetized that all of one face exhibits a polarity of one sign and all of the other face a polarity of the opposite sign, we have a magnetic shell. Such a magnetized sheet, or lamina, may be thought of as made up of a large number of very short magnets (dipoles) all having like poles facing in one direction. In dealing with such a laminal magnet, it is considered that the material constituting the shell is magnetized at each point in a direction normal to the face at that point.

What is known as the **strength** of such a magnetic shell, at any point, is given by the product of the intensity of magnetization and the thickness of the shell at that point. Thus

$$S = It, \tag{23}$$

where S represents the strength of the shell, I the intensity of magnetization, and t the thickness of the shell. By Eq. (21) this relation may be given the form

$$S = \frac{M}{V} t = \frac{M}{a} \tag{24}$$

where M is the magnetic moment and a the area of one face of the lamina. This means that **the strength of the shell is numerically equal to the magnetic moment per unit area.** The concept of a magnetic shell will be found useful when we come to deal with the magnetic effects produced by electric currents.

26. Field Intensity at a Point on the Major Axis of a Magnet. Now that we have considered the concept of magnetic moment we are in a

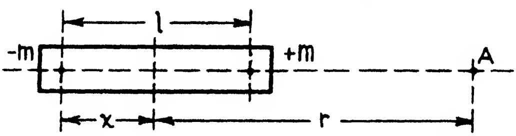


FIG. 14.—Field intensity at a point on the major axis of a magnet—the Gaussian A position.

position to return to the subject discussed in Sec. 20. There we dealt with the idea of intensity at **any** point in a magnetic field. There are **two particular directions** along which it is important to be able to compute the field intensity. One such direction is indicated in the sketch constituting Fig. 14. In his researches on magnetism, Gauss computed the field intensity at various points in the region of a magnet. He used the letter A to indicate a point on the major axis of the magnet and B to

represent a point on a median line at right angles to the principal axis. Accordingly, the case diagrammed in the figure is known as the **Gauss A position.** Let A, then, be the point at which we wish to determine the field intensity. The algebraic work will be lessened if we use the half-length x instead of the total distance l between the pole centers.

The resultant field at A will be the vector sum of the fields due to the poles $+m$ and $-m$; this sum may be written

$$H_A = H_{+m} - H_{-m}.$$

Assuming the medium to be air, and applying Coulomb's law [Eq. (17)], we rewrite the above equation as follows

$$H_A = \frac{m}{(r-x)^2} - \frac{m}{(r+x)^2} = \frac{4rxm}{(r^2-x^2)^2} = \frac{2r(2xm)}{(r^2-x^2)^2}.$$

But $2xm = lm = M$, the moment of the magnet; hence

$$H_A = \frac{2rM}{(r^2-x^2)^2} \qquad \text{oersteds.} \tag{25}$$

If A is far enough from the magnet so that x^2 is very small compared to d^2, then the factor x^2 may be neglected and Eq. (25) becomes

$$H_A = \frac{2M}{r^3}. \tag{26}$$

If x is $r/10$, x^2 will be only 1 per cent of r^2. If greater accuracy is desired, Eq. (25) should obviously be used. It is also to be noted that the distances involved are measured to the pole centers, the location of which is somewhat uncertain. Unless the poles can be definitely located, a certain amount of error in H will always be involved.

27. Field Intensity at a Point on a Median Line Perpendicular to the Major Axis. This is the Gauss B position, as shown in Fig. 15. In this case our problem is to derive an expression which will give the magnetic field intensity at a point B equidistant from the poles of the magnet.

The intensity H_N at B due to the N pole will be m/d^2, directed toward D. The intensity H_S due to the S pole will be m/d^2, directed toward C. The resultant intensity H will be the vector sum of H_N and H_S, directed toward R. On the assumption that $+m$ and $-m$ are equal, H_N and H_S will be equal, and hence BR will be parallel to the magnet. Accordingly, NBS and BDR are similar triangles, and we may write

$$\frac{l}{d} = \frac{H_B}{H_N},$$

or

$$H_B = \frac{lH_N}{d}.$$

Substituting the equivalent values for H_N, d, and l, and simplifying, we have

$$H_B = \frac{2xm}{d^2(x^2 + r^2)^{1/2}} = \frac{2xm}{(x^2 + r^2)^{3/2}}.$$

But $2xm$, or lm, is the magnetic moment, hence

$$H_B = \frac{M}{(x^2 + r^2)^{3/2}} \quad \text{oersteds.} \tag{27}$$

As in the previous case, if r is great compared with x, Eq. (27) reduces to

$$H_B = \frac{M}{r^3} \quad \text{oersteds.} \tag{28}$$

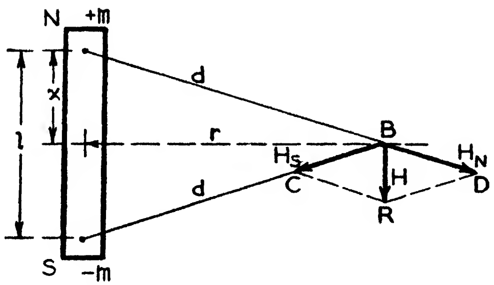


FIG. 15.—Field intensity at a point on a line perpendicular to the major axis—the Gaussian B position.

A comparison of Eqs. (26) and (28) shows that at points considerably removed from the magnet the field intensity at a point in line with the axis is twice what it is at a corresponding distance on a median line perpendicular to the magnet, but in both cases the magnitude of the resultant field varies inversely as the **third** power of the distance.

Problem. A magnet has a pole strength of 100 units and its magnetic length is 10 cm. What is the magnitude of the field strength at a point on its major axis 25 cm from its N pole?

Solution. Make use of Eq. (25) and we have

$$H_A = \frac{2 \times 30 \times 10 \times 100}{(30^2 - 5^2)^2} = 0.0622 \text{ oersted.}$$

The approximate formula [Eq. (26)] yields

$$H_A = \frac{2 \times 10 \times 100}{30^3} = 0.0742 \text{ oersted.}$$

Thus we see that in this case the simplified formula gives a result which is nearly 20 per cent too large. Equation (26) is therefore not applicable **under these circumstances.**

28. Comparison of Fields by the Oscillations of a Magnet. The foregoing discussions indicate that the mechanical interaction between the poles of a magnet and any magnetic field in which the magnet may be located depends upon (1) the magnetic moment M, and (2) the intensity of the surrounding field. It therefore becomes possible to compare the intensity of two fields by observing the mechanical behavior of the magnet when suspended in the two fields. To carry out this procedure a magnet is supported in such a way that it is free to oscillate about a vertical axis—say by being suspended by means of a light untwisted thread.

Suppose any bar magnet, for example, is supported in a field of known intensity, as indicated above. If, after the magnet has come to rest, it be mechanically deflected from its position of equilibrium through a small angle θ and then set free, it will oscillate about the axis of support. The restoring couple will be supplied by the surrounding field. We have already seen (Sec. 23) that the magnitude of this restoring torque is given by the relation

$$L = MH \sin \theta; \tag{i}$$

and if $\sin \theta$ is small, this equation may take the form

$$L = MH\theta. \tag{ii}$$

This torque will give to the magnet an angular acceleration indicated by the expression

$$L = -I\alpha, \tag{iii}$$

or

$$\alpha = -\frac{L}{I}, \tag{iv}$$

where I is the moment of inertia of the magnet and α the angular acceleration. Obviously the angular acceleration is proportional to the displacement and is oppositely directed; hence the negative sign in (iii). Thus the basic conditions for simple harmonic motion are fulfilled. We may equate the value of the torque as given in (ii) and (iii). This yields

$$MH\theta = -I\alpha.$$

Rearranging terms

$$\frac{I}{MH} = -\frac{\theta}{\alpha}. \tag{v}$$

Since the angular motion is harmonic, the period will be given by

$$T = 2\pi\sqrt{-\frac{\theta}{\alpha}}. \tag{vi}$$

Substituting in (vi) the values as given by (v), we have

$$T = 2\pi\sqrt{\frac{I}{MH}}, \tag{29}$$

or

$$H = 4\pi^2\left(\frac{I}{MT^2}\right). \tag{30}$$

Thus we have a relation that expresses the field intensity H as a function of the magnetic moment of the magnet, its moment of inertia, and its period of angular oscillation. For a given magnet, the case may be expressed thus,

$$H = A\frac{1}{T^2} = An^2 \tag{31}$$

where A is the constant numerically equal to $4\pi^2 I/M$ and n the frequency of oscillation.

If then one desires to compare the intensity of two different magnetic fields, either the period or the frequency of oscillation of a given magnet may be noted, when the magnet is supported by the same suspension in the two fields. From Eq. (31) we may set up the working relation

$$\frac{H_1}{H_2} = \frac{T_2^2}{T_1^2} = \frac{n_1^2}{n_2^2}, \tag{32}$$

and substitution in this equation will give us the comparison desired. If the field intensity in one of the cases is known, the magnitude in the second case can be computed. In practice a cylindrical or rectangular bar magnet is usually employed, and is suspended in such a manner that air currents may be avoided.

29. Determination of Absolute Values of M and H. In Eq. (30) we already have a relation involving magnetic moment and field intensity. If, then, we can derive another expression in which these two quantities appear it will be possible to eliminate either one of these terms and thus evaluate the other. Equation (30) may be rewritten as

$$MH = \frac{4\pi^2 I}{T^2}, \tag{33}$$

thus making it possible to find the product of M and H once the moment of inertia and the period are known. I may be computed from the physical dimensions of the magnet used, and its period can be determined easily by experimental means.

In order to find a second relation between M and H, use is made of the so-called **tangent law.** If and when external magnetic fields, at right angles to one another, are caused to act on a magnet, the magnet will tend to set itself parallel to the resultant field. The horizontal component of the earth's field (Sec. 37) will serve as one of these two fields, and a bar magnet, so positioned that its major axis is at right angles to the earth's field, will serve to provide the second field needed. The situation is

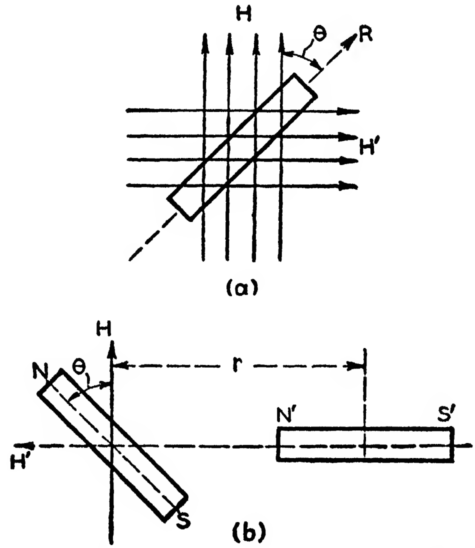


FIG. 16.—Resultant magnetic field. (*a*) Position of a magnet when subjected to two mutually perpendicular fields; (*b*) magnetometer needle and auxiliary deflecting magnet.

shown, diagrammatically, in Fig. 16*a*. In the sketch, H represents the earth's field and H' the field due to a bar magnet. A compass needle, or a delicately suspended small magnet, will take up a position along R parallel to the resultant field. When equilibrium is attained the suspended magnet will make an angle θ with the earth's field and an angle $90 - \theta$ with H'. The moment of the couple tending to rotate the magnet in the direction of H would be $MH \sin \theta$, and that tending to rotate it in the direction of H' would be $MH' \sin (90 - \theta)$, or $MH' \cos \theta$. When mechanical equilibrium obtains, these two torques are equal, and we may therefore write

$$MH \sin \theta = MH' \cos \theta.$$

This is equivalent to

$$\frac{H'}{H} = \frac{\sin \theta}{\cos \theta} = \tan \theta. \tag{34}$$

The arrangement of the components that is used to secure data from which to solve Eq. (34) is sketched in Fig. 16b. NS is the suspended magnet (usually housed), the assembly being known as a **magnetometer**; $N'S'$ is the magnet whose MH value has previously been determined.

In order to eliminate H', use may be made of the relation expressed by Eq. (26). Such a step is justified because the setup shown in Fig. 16b is essentially the Gaussian A position. If, then, the magnet $N'S'$ is so placed that $r \gg l/_2$, H' will equal $2M/r^3$. Substituting this for H' in Eq. (34) and rearranging terms, we have

$$\frac{M}{H} = \frac{r^3 \tan \theta}{2}, \tag{35}$$

which gives us the desired **second** relation between M and H. If r is not great in comparison to $l/2$, it will be necessary to substitute the more exact value for H' as given by Eq. (25). The student should make that substitution.

By combining Eqs. (33) and (35) we may secure the working relations

$$M = \left(\frac{2\pi^2 I r^3 \tan \theta}{T_1^2}\right)^{\frac{1}{2}} \tag{36}$$

and

$$H = \left(\frac{8\pi^2 I}{T^2 r^3 \tan \theta}\right)^{\frac{1}{2}}. \tag{37}$$

Thus we have a relation by means of which the magnetic moment of a magnet may be computed, and another equation by which the absolute value of field intensity may be obtained. It is also to be noted that the expressions for M and H involve only quantities which may be readily determined experimentally.

30. Magnetic Flux. In dealing with electrostatic induction it was found that a relation existed between electrostatic field strength and electrostatic flux, or displacement (Sec. 8). In that case, also, it was noted that any medium in the field was in a state of distortion (strain) and that this held for "free" space. A corresponding condition obtains in a magnetic field. Magnetic flux is related to field intensity by the equality

$$B = \mu H, \tag{38}$$

where B is the magnetic density (lines of flux per unit area), μ the permeability, and H the field strength (sometimes incorrectly called **magnetizing force**). But Eq. (17) indicated that

$$H = \frac{m}{\mu d^2};$$

hence

$$B = \frac{m}{d^2}, \tag{39}$$

Thus B (commonly referred to as **magnetic induction**) depends upon pole strength m and distance d, and **not upon the medium.** If the medium is air (strictly, free space) μ will be unity, and hence H and B will be given numerically by the same expression, viz. m/d^2. Magnetic induction (flux density) is expressed in terms of lines per unit area, or in **gausses.** Magnetic intensity and magnetic induction should not be confused; the

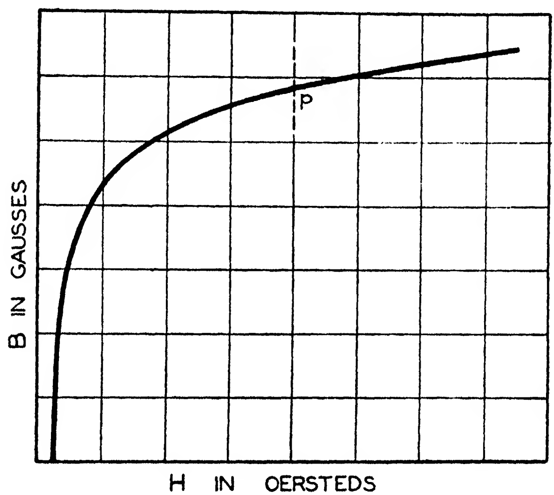


FIG. 17.—Relation between magnetizing field and the resulting induction.

former has to do with the **property of the field** expressed in terms of force, the latter with the **condition** of the medium. Magnetic induction is a vector quantity; its direction is that of H.

Referring again to the relation indicated by Eq. (38) it should be noted that, since permeability μ is not constant (it depends upon the magnitude of the field intensity H), the induction B is not a linear function of H. For many magnetic materials, the graphic representation of this relation will be indicated by a curve such as that sketched in Fig. 17. It will be noted that, as the field intensity is gradually increased in magnitude, a condition is reached (in the region of p) where an increase in H does not produce an increase in the induction (flux density). The material is then said to be **magnetically saturated.** The explanation for this behavior is that the dipole entities have probably all become magnetically aligned with the exciting field. The curve shown in Fig. 17 is a highly

important graph and is commonly referred to as the *B-H* curve. Its exact contour will depend upon the particular material being tested. We shall return to this relation later.

In dealing with the magnetic flux through some given area, we are usually interested in considering the **normal component** of that flux. On that basis magnetic flux may be defined by the relation

$$\Phi = \int_s B \, da \cos \theta, \tag{40}$$

where θ is the angle between the normal to the surface da and the actual direction of the lines of induction. If and when the field is uniform, and the direction of the induction is normal to the surface, Eq. (40) becomes

$$\Phi = BA, \tag{41}$$

where Φ is the total number of lines of flux and A the area involved. When dealing with **total flux** the unit is the **maxwell.** A maxwell is equivalent to one line of magnetic flux.

The total flux emanating from unit pole is also of special interest. If we imagine a sphere of 1 cm radius surrounding a unit pole, the total flux can be readily computed. The flux density at any and all points on the surface of such a hypothetical sphere would be unity, and would be represented by one line per square centimeter. The area of the sphere would be 4π cm². By Eq. (41) the total flux emanating from unit pole would therefore be 4π maxwells. For any pole strength other than unity, the total flux issuing would be given by the expression $4\pi m$, where m is the pole strength in question.

If the induction is not normal to the surface, we may apply Gauss's basic theorem to the magnetic case and arrive at the conclusion that the total normal induction (total normal flux) over any closed surface drawn about a pole is given by the expression

$$\Phi = 4\pi m \qquad \text{maxwells.} \tag{42}$$

It may be shown that if the pole is outside the surface the total normal flux over the surface is zero.

Problem. A pole whose strength is 200 units is located at the center of a sphere having a diameter of 4 cm. What is the flux density at the surface of the sphere, and what is the total flux?

Solution. By Eq. (39), the flux density will be

$$B = \frac{200}{2^2} = 50 \text{ gausses.}$$

The total flux, by Eq. (42), will be

$$\Phi = 4\pi 200 = 800\pi \text{ maxwells.}$$

31. Force between Two Magnetized Surfaces. One important application of the principles discussed in the preceding section has to do with the determination of the magnitude of the mechanical force exerted by one magnetized surface upon another similar surface, particularly when the surfaces are close together; one familiar example being the force exerted by a permanent magnet or an electromagnet on a soft iron body. Let us develop a relation that will give the magnitude of the force in such a case.

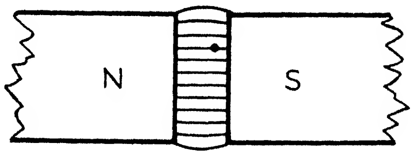


FIG. 18.—Magnetic field in the region between two unlike poles.

If the distance of separation of the two magnetized bodies is small in comparison with the area (Fig. 18), the lines of flux will be parallel except very near the edges. We will assume the field to be uniform. Let the intensity of magnetization be I and the area of each opposing surface be A. It has been shown (Sec. 30) that 4π lines of flux leave each unit pole, and since there are I such poles per unit area, it follows that the flux density (induction) in the air gap will be $4\pi I$ gausses.

Since the medium is air, $B = H$; therefore the force experienced by unit test pole N placed anywhere (except near the edges) in the air gap will be equal to $4\pi I$ dynes. Half of this force $2\pi I$ will be due to the attraction of the S pole and the other half to the repulsion of the N pole. Since each unit area of a given pole, say the N pole, exerts a force of $2\pi I$ dynes on a test pole anywhere in the air gap, and since I pole units exist on each unit area of the S pole, the force experienced by that area (1 cm²) will be equal to $2\pi I^2$ dynes. The total force, therefore, which the N pole will exert on the S pole will be

$$F = 2\pi I^2 A,$$

where A is the area of the opposing magnetized surfaces. As already indicated, in this case

$$B = H = 4\pi I.$$

Substituting the value for I obtained from this equation, in the above expression for force we get

$$F = \frac{B^2 A}{8\pi} \quad \text{dynes,} \tag{43}$$

where B is the flux density in gausses and A the area in square centimeters. If A is expressed in square inches, and B in lines per square inch, Eq. (43) becomes

$$F = \frac{B^2 A}{72{,}130{,}000} \quad \text{lb.} \tag{44}$$

It will be noted that the distance of separation does not enter into the above expressions for force. However, as the bodies are moved farther away, more flux will spread into the region beyond the edges and thus contribute less to the resulting force; hence F will tend to diminish somewhat as the bodies are moved apart.

Problem. One pole of a magnet measures 1×1 cm. A piece of iron having an equal area is placed close to the pole face. A total uniform flux of 50,000 maxwells passes from the pole face to the piece of iron. What force of attraction does the magnet exert upon the iron? Express the result in grams and in pounds. Let $g = 980$ cm/sec^2.

Solution. By Eq. (43), we have

$$\begin{aligned} F &= \frac{(50{,}000)^2 \times 1}{8\pi 980} = 1.018 \times 10^5 \text{ gm} \\ &= 10.18 \text{ kg} \\ &= 22.5 \text{ lb.} \end{aligned}$$

32. Energy per Unit Volume of the Medium. It has already been indicated that the medium (including free space) in which a magnetic field obtains is in a state of strain. This implies that energy has been expended to bring about that condition. A magnetic field must therefore involve an energy content; and this potential energy plays a very important part in many electromagnetic processes which we shall examine later.

Following the line of reasoning that we employed in connection with the corresponding case in electrostatics (Sec. 16), we may make use of the force relation that was developed in the preceding section.

From Eq. (43) we know that the force in dynes of the attraction between two magnetized surfaces, 1 cm^2 in area, is given by

$$F = \frac{B^2}{8\pi}.$$

Referring to the situation shown in Fig. 18, let us assume an area 1 cm^2, no part of which is near an edge. If we move the magnetized bodies involving the assumed area slightly farther apart, say a distance of dx, the work done in overcoming the above indicated force will be

$$dw = \frac{B^2}{8\pi}\, dx.$$

If these magnetized surfaces are moved 1 cm apart, thus increasing the volume by unity, the work done would be

$$\int dw = \frac{B^2}{8\pi} \int_{x=0}^{x=1} dx,$$

or

$$\text{Total work} = \text{P.E.} = \frac{B^2}{8\pi} \qquad \text{ergs/cm}^3. \tag{45}$$

Thus each centimeter of a uniform magnetic field in free space has an energy content given by the above expression. If the region in which the field exists has a permeability other than unity, μ will appear in the numerator of the right-hand side of Eq. (45).

Problem. Referring to the numerical problem given in connection with the discussion in Sec. 31, what would be the energy content of 1 cm³ in that case?

Solution. Substitution in Eq. (45) yields

$$\text{P.E.} = \frac{(50{,}000)^2}{8\pi} = 1.018 \times 10^5 \text{ ergs.}$$

33. Refraction of Lines of Magnetic Flux. The direction of magnetic flux is changed when it crosses the boundary between two mediums having different permeabilities. The line of reasoning followed in dealing

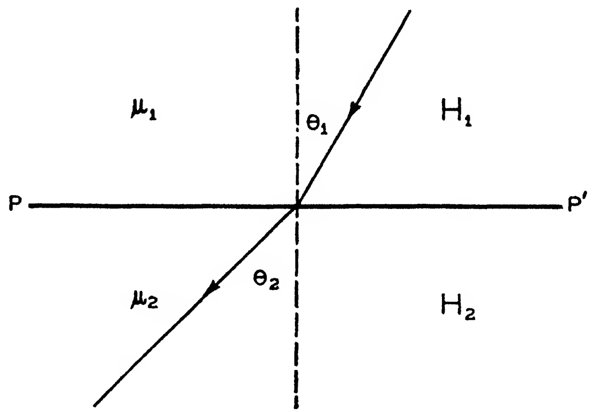


FIG. 19.—Refraction of lines of magnetic flux.

with the corresponding case in electrostatics (Sec. 17) is applicable here. The situation is sketched in Fig. 19. PP' represents the boundary between the two mediums, whose permeabilities are μ_1 and μ_2 respectively. As in the electrostatic case, two conditions must be satisfied if the flux lines are to be continuous: (1) the **tangential** component of the **magnetic intensity** (H) in the first medium must equal the tangential component of the intensity in the second medium, and (2) the **normal** component of the **induction** (B) in the first medium must equal the corresponding component in the second medium. Accordingly, it follows that

$$\frac{\tan\,\theta_1}{\tan\,\theta_2} = \frac{\mu_1}{\mu_2}. \tag{46}$$

If, then, the medium above PP' was air ($\mu = 1$), and the medium below of such a nature that its permeability was **greater** than unity, the flux would be refracted **away from the normal,** *i.e.*, θ_2 would be greater than θ_1. Such a case is illustrated in Fig. 20. Here we have a field in air in which is located, say, a soft-iron cylindrical core. Due to the principle embodied in Eq. (46), the flux will be refracted at the boundary with the result that the flux is concentrated in the iron core. Use is made of this possibility in the design of galvanometers and portable electrical-measuring instruments, as we shall see later.

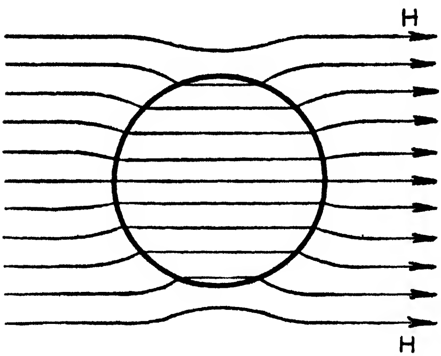


FIG. 20.—Showing the refraction of magnetic flux due to the presence of a body of magnetic material in the field.

The reverse case is also important. While there is no known "insulator" for magnetic flux, it can be diverted, in part at least, from a specified region by taking advantage of the relation embodied in Eq. (46). Such a case is sketched in Fig. 21. It frequently happens that one desires to screen a given region from magnetic flux. This can be done by enclosing the region with a medium, such as soft iron, whose permeability is relatively high. An iron cylinder C will serve to refract the flux around the space S. By this procedure galvanometers and other electrical-measuring instruments may be shielded from the effect of the earth's magnetic field, and also from the stray fields due to other sources, such as electromagnets.

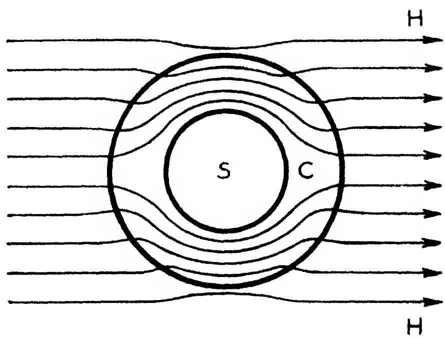


FIG. 21.—Showing the screening effect due to the presence in the field of a material having relatively high permeability. Note that no flux exists in the region S.

34. Induced Magnetism. We may sum up our discussion of magnetostatics by briefly considering the general relations which exist between several of the more important concepts which we have examined in the preceding sections.

If a piece of magnetizable material, such as a piece of soft iron, is brought into a region where a magnetic field exists, the atomic or molecular

dipoles will orient themselves in such a manner that each end of the piece will manifest polarity as long as the body remains in the field.[1] In any given case the intensity of magnetization (Sec. 22) developed will depend for its magnitude on the field intensity, as given by the expression

$$I = \kappa H. \tag{47}$$

The proportionality factor κ is known as the **susceptibility** of the material. By this is meant the extent to which magnetization may be developed in any given kind of material. A quantitative definition of susceptibility may at once be had from Eq. (47), thus

$$\kappa = \frac{I}{H}, \tag{48}$$

i.e., it is defined as the ratio of the intensity of magnetization to the field intensity. This factor has a relatively large numerical value in the case of such substances as iron, cobalt, nickel, etc., but it may even be **negative** as in the case of such materials as bismuth. Like permeability (μ), susceptibility in the case of certain substances is not constant but is found to be a function of the temperature and of the magnitude of the magnetizing field H.

A second important relation between H and I may be easily deduced. It has already been shown (Sec. 31) that at any point near a magnetized surface the flux density in air is given by the relation

$$H = 4\pi I. \tag{49}$$

If and when a piece of magnetizable material occupies a position in a field, the flux within the region occupied by the material will consist of two components, viz., the flux H constituting the magnetizing field plus the flux, say H', induced in the material. This relation could be expressed thus

$$B = H + H'.$$

But H' is identical with the field value expressed by Eq. (49) because each of the lines issuing from the body must pass through the substance. It accordingly follows that

$$B = H + 4\pi I. \tag{50}$$

We shall find this to be an important relation when we come to study the magnetizing effects of the electric current.

From the above equality it is possible to deduce a relation between permeability μ and susceptibility κ. By substituting in Eq. (50) the

[1] This subject will be more fully discussed in connection with electromagnetism.

value of I given in Eq. (47), there results

$$B = H + 4\pi\kappa H = (1 + 4\pi\kappa)H. \tag{51}$$

But we have already seen [Eq. (38)] that

$$B = \mu H;$$

hence it follows that

$$\mu = 1 + 4\pi\kappa. \tag{52}$$

The above expression gives us another definition of permeability.

It should be pointed out that Eq. (50) does not quite tell the whole story when dealing with induced magnetism. In the first place the magnetization may not be entirely uniform throughout the substance involved; and secondly, there is a certain **demagnetizing effect** due to the induced poles themselves. If a soft iron core is placed within a helix through which a current is passing the core will be magnetized, and the lines of flux extending from pole to pole of the core, **outside** of the bar, will be in a direction **opposite** to the inducing field. The resultant field will therefore be **less** than the value of H given by the above equations. Thus the effectiveness of H is reduced, and by an amount which depends upon the strength of the existing poles.

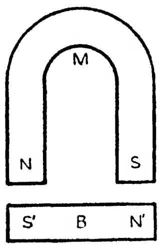


FIG. 22.—Horseshoe form of magnet, with soft iron bar joining the poles.

This demagnetizing effect obtains in the case of permanent magnets as well as in a piece of material in which magnetism is being induced by an external field. It is for this reason that bars, or "keepers," are often placed across the poles of a permanent magnet. The situation is illustrated in Fig. 22. If, in the case of the familiar horseshoe magnet shown, we can nearly close the gap by means of a soft iron bar B, this bar will become magnetized by induction and will become polarized as indicated. Accordingly, the poles of this bar, N' and S', will tend to neutralize the demagnetizing fields due to N and S. Thus the demagnetizing tendency will be materially lessened, and the permanent magnet will maintain its pole strength for a longer period of time.

All dielectrics behave in much the same way when placed in an electrostatic field, but all materials do not react alike when subjected to a magnetic field. For example, one encounters certain substances whose susceptibility is somewhat greater than unity and whose permeability is correspondingly in excess of unity—something of the order of 1 per cent. Such materials are referred to as **paramagnetic substances.** Examples of this type of body are aluminum, manganese, titanium, and liquid oxygen.

There is a second group of substances that show a high degree of susceptibility and a correspondingly great permeability, in some cases reaching a μ value of several hundred. Among these may be mentioned iron, cobalt, nickel, and certain magnetic alloys. Such materials are known as **ferromagnetic substances.**

Magnetic alloys are becoming increasingly important. The first important magnetic alloy to be produced was that developed by Heusler, some 40 years ago. The Heusler alloy contains no iron; it is composed of 14.3 per cent aluminum, 28.6 per cent magnanese, and 57.1 per cent copper. In this alloy the permeability increases with the addition of manganese up to the point where the proportion of manganese and aluminum are in proportion to their atomic weight. The maximum magnetization which may be obtained with this alloy is about one-third of that of the best iron. The copper which enters into its composition apparently does not affect its magnetic properties, serving only to make the alloy more easily machined.

Recently, other magnetic alloys have been developed which find wide use in connection with power and communication engineering. These alloys are composed largely of iron and nickel in varying proportions. They possess several important magnetic characteristics which will be touched upon in connection with our later study of electromagnetism.

A third group of bodies known as **diamagnetic substances** show **negative** susceptibility, and a permeability slightly less than unity. Such substances tend to move into the weakest part of a magnetic field. Bismuth, whose κ value at room temperature is approximately -1.5×10^{-6} and whose μ value is about 0.9983, is one of the few examples of this type of substances. No particular use has thus far been made of the peculiar magnetic behavior of diamagnetic bodies, but their existence does enter into the problem of accounting for magnetic phenomena.

35. Permanent Magnetism. Certain substances, when subjected to a magnetizing field, retain a substantial amount of magnetization after the magnetizing force has been removed. It thus becomes possible to produce permanent magnets. Various kinds of steel exhibit this useful property to a marked degree. Steel, as is well known, consists of iron to which has been added a small, but definite, amount of carbon, the whole having been subjected to a special heat treatment. It would appear that the atomic dipoles in steel, when once magnetically oriented, do not, of themselves, readily return to their random arrangement. Why this is so has not yet been fully explained. Various other elements instead of carbon are used as the hardening component, notable among which are tungsten, cobalt, chromium, silicon, and manganese. Tungsten and cobalt steel are particularly useful as material for permanent magnets,

especially the latter. Recently, a special permanent-magnet type of steel, going under the trade name of **Alnico,** has been produced which is particularly useful for certain types of magnets. This alloy consists of iron, nickel, and aluminum. Such magnetic alloys are subjected to special and closely controlled heat treatment when being formed, and also during the subsequent process of magnetization. The extent to which these metallurgical and magnetizing techniques have been developed is attested by the permanence of the calibration of electrical-measuring instruments of which permanent magnets form a part.

When dealing with permanent-magnet material, the magnetization which remains **after** the magnetizing field is reduced to zero is called **remanence.** When the magnetic material has been saturated and the magnetizing field reduced to zero, the **induction** which remains is referred to as the **retentivity** of the material. In certain of the special steels above mentioned, both the remanence and retentivity are high. These concepts will enter into our study in a later chapter.

36. Temperature Effects. Reference has been made to the part played by temperature in connection with several magnetic phenomena. One or two statements should perhaps be added to the observations already made.

When a sample of ferromagnetic material, such as iron, is subjected to an increasing temperature, it tends to lose its ability to become magnetic and, at a rather definite temperature, it abruptly ceases to be ferromagnetic and becomes paramagnetic in character. The temperature at which this change occurs depends upon the particular specimen being examined. For specimens of iron and steel the change occurs between 700 and 900°C. Hopkinson, an early investigator in this field, called this transition temperature the **critical temperature;** it is often referred to as the **Curie point,** after another research worker. The exact form of the curve relating permeability and temperature depends upon the intensity of the magnetizing field at which the temperature test is carried out. This is strikingly shown by the several graphs appearing in Fig. 23; these curves are for pure iron. In this case the transition occurs at a temperature slightly below 800°C. This is in the temperature region commonly designated by the term "dull-red heat." In the case of nickel the critical temperature lies between 310 and 350°C; for cobalt it is in the region of 1150°C. If a permanent magnet is heated to the Curie temperature (787.5°C), it loses its magnetic properties and does not, of itself, regain them on being cooled to ordinary temperatures.

It would appear that some marked atomic or molecular change occurs in a magnetic material at or near the critical temperature. For instance, in the case of iron the Curie point coincides with the temperature at

which one encounters rapid changes in such physical characteristics as density, specific heat, and electrical resistance. Indeed, the critical temperature is the same temperature at which recalescence[1] occurs. Though a considerable amount of research has been carried out in this

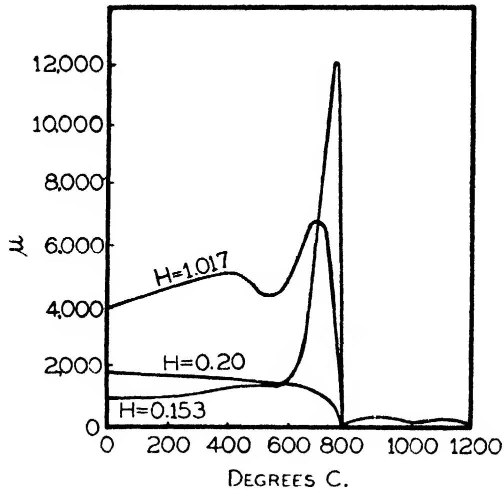
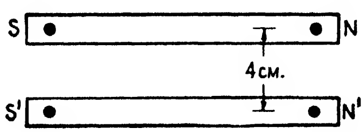


Fig. 23.—Showing the effect of heat on the magnetic permeability of iron, when subjected to different magnetizing fields. *(From "Magnetic Phenomena" by S. R. Williams.)*

field, we have as yet no thoroughly satisfactory explanation for the existence of these important concomitant phenomena. Here is a field which might well serve as a challenge to the efforts of any young and ambitious investigator.

PROBLEMS

1. Two bar magnets are placed parallel to one another as shown. If the pole strength in each case is 100 units and the length 10 cm, what will be the magnitude of the force tending to push the magnets apart?



2. In a bar magnet the distance between the poles is 10 cm, and the pole strength is 100 units. What is the field intensity at a point 5 cm from the N pole on a line at right angles to the major axis of the magnet? What direction will the resultant force make with reference to the axis of the magnet?

[1] When a piece of steel is allowed to cool rapidly from a bright red heat, it will, at a certain well-defined temperature, suddenly become brighter and emit an unusual amount of heat.

3. In Prob. 2, what would be the field intensity at the same distance from the same pole, but on the line of the axis extended? If a pole of strength 50 units were placed at this point, what would be the magnitude and direction of the force to which it would be subjected?

4. The distance between the poles of a magnet is 8 cm; the strength of each pole is 100 units. This magnet is suspended in a horizontal position in a field of 0.2 oersted. What will be the magnitude of the torque tending to restore it to its normal position of equilibrium when it is deflected through an angle of 45° by the application of a horizontal mechanical torque?

5. Prove that the field intensity at a point near a large sheet of uniformly magnetized material is equal to 2π times the intensity of magnetization of the surface.

6. The field intensity at a certain point is known to be 0.18 oersted. At that point the period of a certain horizontally-suspended magnet is 12 sec. The same magnet, supported by the same suspension, when moved to another point in the field is found to have a period of 12.5 sec. What is the value of the field at the second location?

7. A cylindrical bar magnet, whose length is 12 cm and whose diameter is 1 cm, is supported in such a manner that it is free to swing in a horizontal plane. It is found to have a period of 10 sec. What is the magnetic moment of the magnet? On the assumption that the pole centers are located 0.5 cm from the physical ends of the magnet, what is the value of the pole strength?

8. A magnet whose magnetic moment is 1,000, upon being suspended in the usual manner in a magnetic field, is found to have a period of 10 sec. What is the field intensity? The moment of inertia of the magnet being used is 800 gm-cm^2.

9. A cylindrical bar of iron, 5 cm in diameter, when placed in a magnetic field of 100 oersteds, is found to have a flux of 62,500 maxwells. What is the permeability of the sample?

10. The core of an electromagnet is 5 cm in diameter. When the magnetizing current is on it is found by experiment that a force of 150 lb is required to separate a piece of soft iron of equal area from a pole of the iron core. What induction exists in the core?

11. If the iron core in Prob. 10 is 6 in. in length what would be the magnetic energy content of the iron?

12. The direction of a certain magnetic field makes an angle of 45° as it passes from air into a body of iron whose permeability is 400. What will be the direction of the flux in the iron?

13. In Prob. 9, what would be the intensity of magnetization under the circumstances indicated?

14. What would be the susceptibility of the iron referred to in Prob. 9?

15. A bar magnet weighing 10 gm is supported by a bifilar suspension 50 cm in length. A similar magnet is placed in a fixed position beside the suspended unit, parallel to it, and with like poles opposite. It is found that the suspended magnet takes up a position 2 cm from the fixed magnet. Assuming that all four poles have the same strength m, what is the magnitude of m?

16. Deduce Eqs. (25) and (27) from Eq. (19).

CHAPTER IV

TERRESTRIAL MAGNETISM

37. The Earth's Magnetic Field. It is probable that the compass was in use as early as 1200, but the fact that the earth functions as a magnet appears not to have been known prior to 1600. In that year Dr. William Gilbert, English court physician, published his classical work on magnetism entitled "De magnete magneticisque corporibus et de magno magnete tellure physiologia nova." In this epoch-making volume, commonly referred to as "De magnete," Gilbert reviewed the then

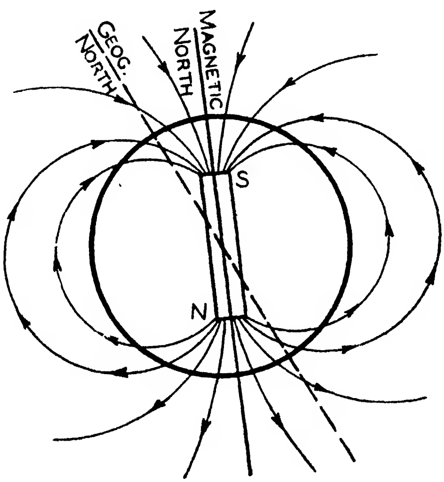


FIG. 24.—Showing the general relation of the earth's magnetic field to the geographical axis.

known information concerning magnetism and added the results of his own careful studies, including his original theory that the earth acts as a great magnet. The science of magnetism began with the appearance of Dr. Gilbert's treatise.

Today we know that the earth acts as a magnet having a north pole located in the vicinity of Boothia Peninsula in Northern Canada, lat. 76°00′ N, long. 102°00′ W, and a south pole in South Victoria Land, lat. 68°12′ S, long. 145°40′ E. These points are not fixed. The terrestrial poles are below the earth's surface. Recent observation indicates that the earth's field extends to considerable distances above the earth's surface. Obviously, the magnetic poles do not coincide with the geo-

graphical poles. The general situation is roughly depicted in the sketch appearing as Fig. 24.

A plane passing through the center of the earth, the axis of support of the compass needle, and Polaris (the north star) would locate the geographical meridian. A plane including the major axis of a compass needle and passing through the center of the earth would give the **magnetic meridian.** The angle between these two planes is referred to as the **declination.** As indicated in the sketch the lines will not be parallel to the earth's surface. In Sec. 22 it was pointed out that a test magnet assumes a position tangential to the direction of the lines of force at the point of observation. It therefore follows that if free to rotate about a horizontal, as well as a vertical axis, a compass needle will not assume a horizontal position. The angle which the needle makes with the horizontal is designated as the **inclination,** or **dip.** The magnitude of the earth's field varies from about 0.18 to something like 0.4 oersted, the latter value occurring in the Philippine Islands. The angle of dip varies from 0 to 90°, being zero at the magnetic equator (see chart, page 59). At the North and the South poles a needle, if free to rotate about a horizontal axis, would assume a vertical position. In dealing with the earth's field it is necessary to resolve the actual field intensity into vertical and horizontal components, as shown in Fig. 25. The horizontal component H acts to turn the compass needle about a vertical axis, and the vertical component V causes the inclination. The ordinary compass needle is mechanically weighted at one end to compensate for its tendency to dip. If the instrument is designed to be used to measure the angle of inclination, it is mounted as shown in the illustration appearing in Fig. 26.

FIG. 25.—The direction of the horizontal and vertical components of the earth's magnetic field relative to the actual direction of the field at any point on the earth's surface.

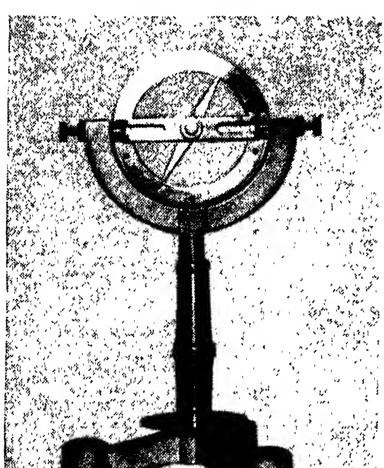


FIG. 26.—Magnetic needle used in determining declination and inclination.

There are various ways of determining the magnitude of the horizontal component of the earth's field and thereby indirectly arriving at the value of the earth's total field. We have already discussed (Sec. 29) the principles involved in one method of determining the horizontal component. In carrying out this determination use is made of what is known as a magnetometer. This apparatus consists of a small magnet suspended in a nonmagnetic housing by a fairly long fiber which is free from twist. The magnet, or the stirrup supporting it, bears a small mirror by means of which any angular deflection may be noted with the aid of the usual telescope and scale. A section of a meterstick extends horizontally in

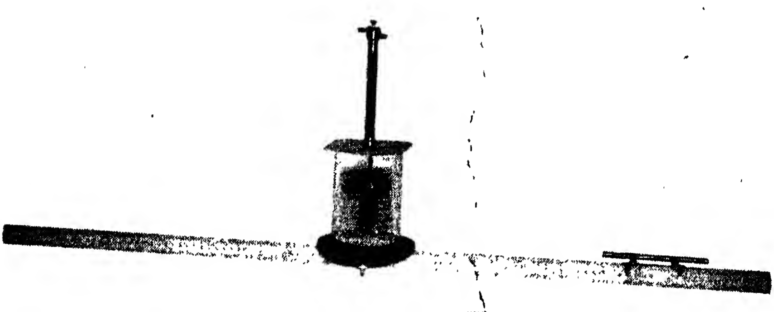


FIG. 27.—Magnetometer assembly. The dark rectangular member within the housing carries the mirror and the small magnet. It also serves as a damping vane.

opposite directions through the axis of support. Some form of sliding clip is provided whereby a second or auxiliary magnet may be supported on, and parallel with, the measuring extensions. The general arrangement of the magnetometer components may be seen in Fig. 27.

The period of the auxiliary magnet and its moment of inertia are first determined. Next the small magnetometer magnet is allowed to come to rest in the magnetic meridian and its exact position noted by means of the optical system provided. The auxiliary magnet is then placed directly on a line at right angles to the magnetometer magnet and with the center distant r centimeters from the axis of the magnetometer needle. By noting the angle through which the magnetometer needle is deflected due to the presence of the auxiliary magnet one is then in a position to make use of Eq. (37) and thereby evaluate H, the intensity of the earth's field.

In carrying out this determination it is necessary to observe certain precautions if accurate results are desired. The magnetometer needle may not be exactly at the center of the scale which carries the auxiliary magnet. To correct for this possible error, deflection readings are taken with the bar magnet in both east and west positions. Also, to compen-

sate for any lack of magnetic symmetry in either the needle or the auxiliary magnet, readings are made in both the east and west positions with the ends of the auxiliary magnet reversed. Similar sets of four readings are made with different values for r.

Another method of determining the magnitude of the earth's field involves the use of a piece of equipment known as an **earth inductor.** This method will be outlined after we have considered the subject of electromagnetic induction.

38. Variation of the Earth's Field. Declination, dip, and the horizontal component of the earth's field intensity are commonly referred to as the **magnetic elements.** As indicated above, the value of the earth's field at any given point is not a fixed quantity; its magnitude and direction are continually changing. Since the earth's field is an important factor (though a diminishing one) in navigation and in civil and electrical engineering, it is important to become acquainted with the character and extent of these changes in the magnetic elements.

The variation in declination appears to have been known to the Italians as early as 1436, and the experience of Columbus on his memorable voyage of discovery clearly established the fact that the declination is radically different at different places on the earth's surface. The first authentic record of magnetic declination was made in the vicinity of London in 1580, the value at that time and place being 11°15′ E. In 1669 the declination there had become zero. Later it swung to the west, reaching a value of 24½° W in 1823, and is now decreasing. From the data secured at the famous Kew Magnetic Observatory and at other magnetic stations throughout the world, it appears that there is a cyclic change in the magnetic declination which has a period of something like 960 years. This type of change is known as a **secular variation.**

In addition to the change just mentioned, the declination is subject to a small **annual variation.** It is an interesting and perhaps significant fact that the yearly variation is in opposite directions north and south of the geographical equator.

There is also a slight daily or **diurnal variation** in all of the magnetic elements which can be detected by delicate magnetic instruments.

It has also been established that there is a cyclic change in the magnitude of the diurnal variations. This superimposed cyclic change has a period of something like 11 years and is frequently referred to as the 11-year period. It happens that there is a maximum of sunspots at regular intervals of 11 to 12 years. Whether sunspots are the **cause** of these variations has yet to be established.

In addition to the more or less regular variations already noted, there are irregular and sometimes violent disturbances of the earth's field.

These erratic variations in the earth's magnetic elements are known as **magnetic storms.** Strangely enough, these disturbances are frequently concomitant with sunspots. In fact, observations extending over a period of 1000 years appear to indicate that sunspot activity and the earth's magnetic activity are in some way related. This relation is strikingly depicted by the graphs shown in Fig. 28. In this connection it is interesting to note that under ordinary circumstances the solar magnetic field has a value of the order of 50 oersteds as compared with a value of something like two-tenths of an oersted for the earth's field. The region of the sun covered by a spot may, however, show a field intensity as high as 4,000 oersteds.

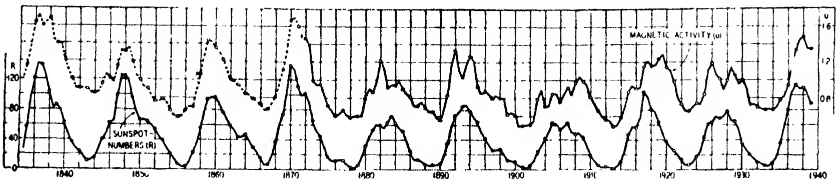


FIG. 28.—Earth's magnetic activity compared with sunspot activity. (*Courtesy Carnegie Institution of Washington.*)

39. Magnetic Maps.—The importance of the study of the earth's magnetic behavior cannot be overemphasized. Through the cooperation of the U.S. Department of Terrestrial Magnetism and the Carnegie Institution of Washington a ship constructed entirely of nonmagnetic materials made several extended voyages for the sole purpose of carrying out magnetic surveys. The U.S. Coast and Geodetic Survey has also collected much valuable magnetic data. Scattered throughout the world are a number of magnetic observatories where continuous observations have been made for many years. Among the older stations may be mentioned the one at Kew already referred to, the one at Potsdam, and another at Bombay. A new station has recently been established in Peru. By the cooperation of these several agencies it has been possible to prepare magnetic maps showing the value of the magnetic elements throughout the world. These constitute what might be called **isomagnetic charts.** If lines are drawn connecting points on the earth's surface having the same declination value we have a chart such as that shown as Fig. 29. Lines of this character are known as **isogonic lines.** It will be noted that the isogonic lines are more or less irregular, and in one instance form a closed loop, the "Siberian oval." There are other isogonals, one passing through North and South America, one through Europe and another through Western Australia, along which the declination is also **zero.** Lines of zero declination are designated as **agonic lines.**

If lines are drawn connecting points showing the same angle of dip we have a map of the character shown as Fig. 30. These lines are referred to as **isoclinic lines.** It will be observed that the isoclinals are consider-

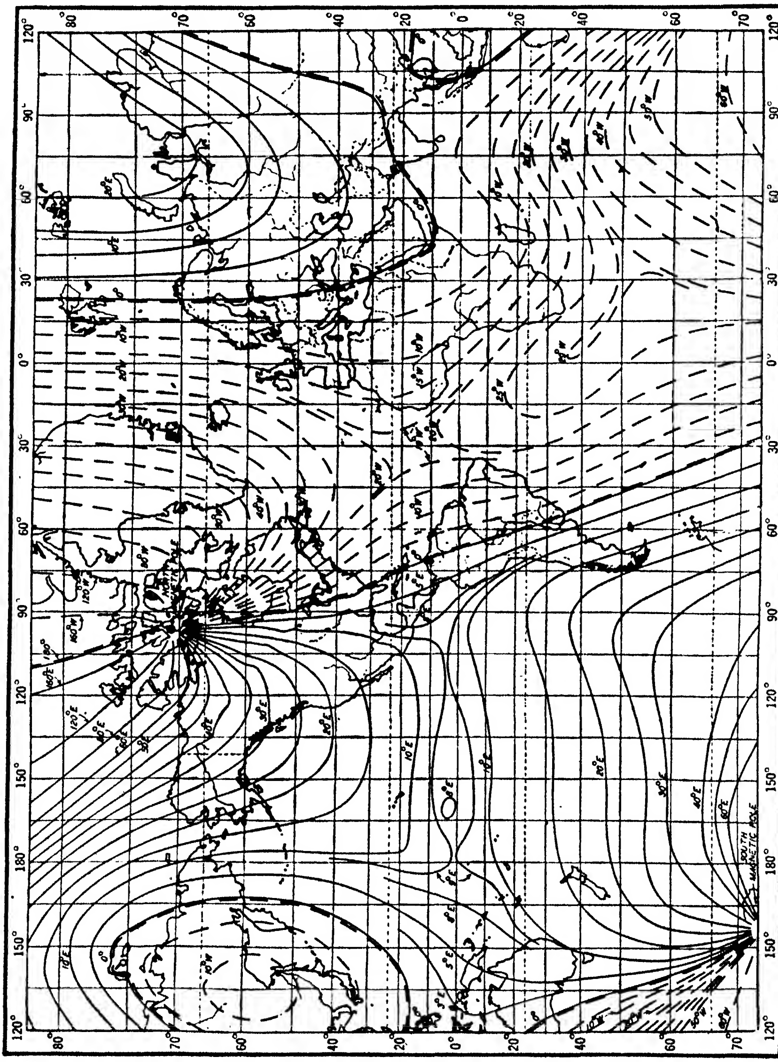


FIG. 29.—Lines of equal magnetic declination (isogonics). (Courtesy of Carnegie Institution of Washington.)

ably more regular that the isogonals, and that they roughly parallel the geographical lines of equal latitude.

Charts are also made (Fig. 31) showing lines which pass through points having the same value for the horizontal component of the earth's magnetic field. Such tracings are referred to as **isodynamic lines.**

40. Theory of the Earth's Magnetic Field. From time to time various theories have been suggested to account for the earth's magnetic field. Anything like a complete explanation of this phenomenon has

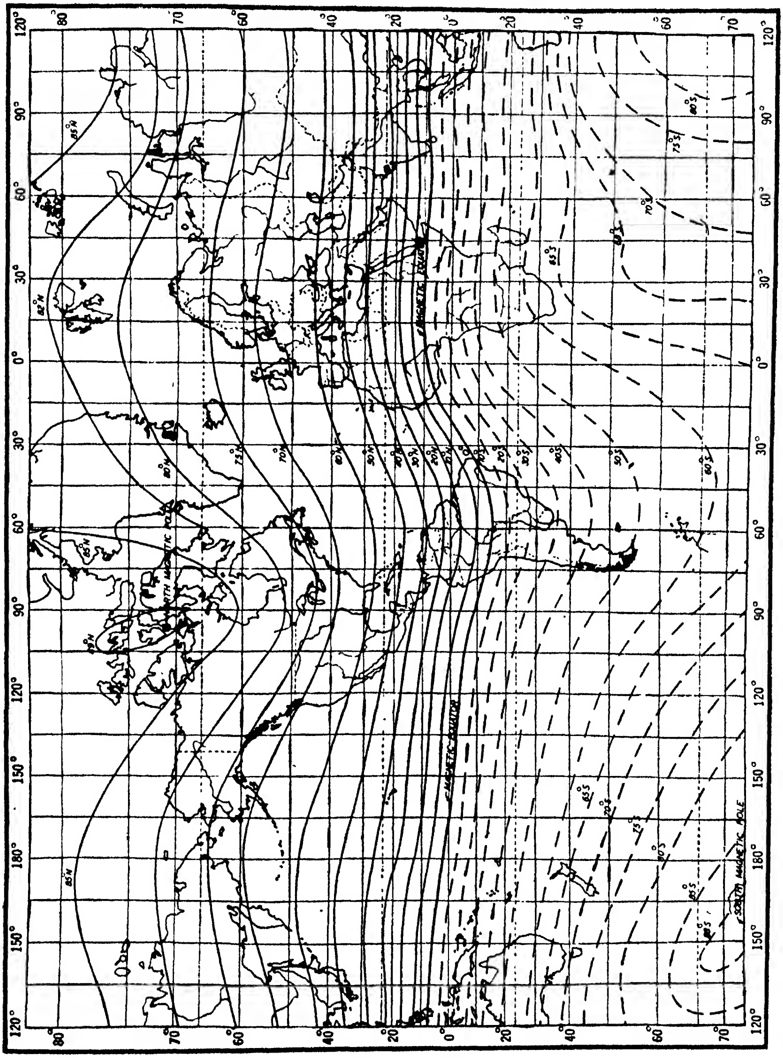


FIG. 30.—Lines of equal magnetic inclination (isoclinics). (*Courtesy of Carnegie Institution of Washington.*)

not yet been offered. Broadly speaking, it may be said that the earth, in some respects, behaves magnetically as if a magnet existed at its center whose longitudinal axis made an angle of about 17° with the axis of the earth (Fig. 24). The long-time changes in declination would appear to

indicate that the axis of such a hypothetical magnet slowly rotates about a fixed point somewhat as does the geographical axis.

It has been shown that the rapid rotation of a magnetizable body will bring about the magnetic state. Since the earth is known to contain a

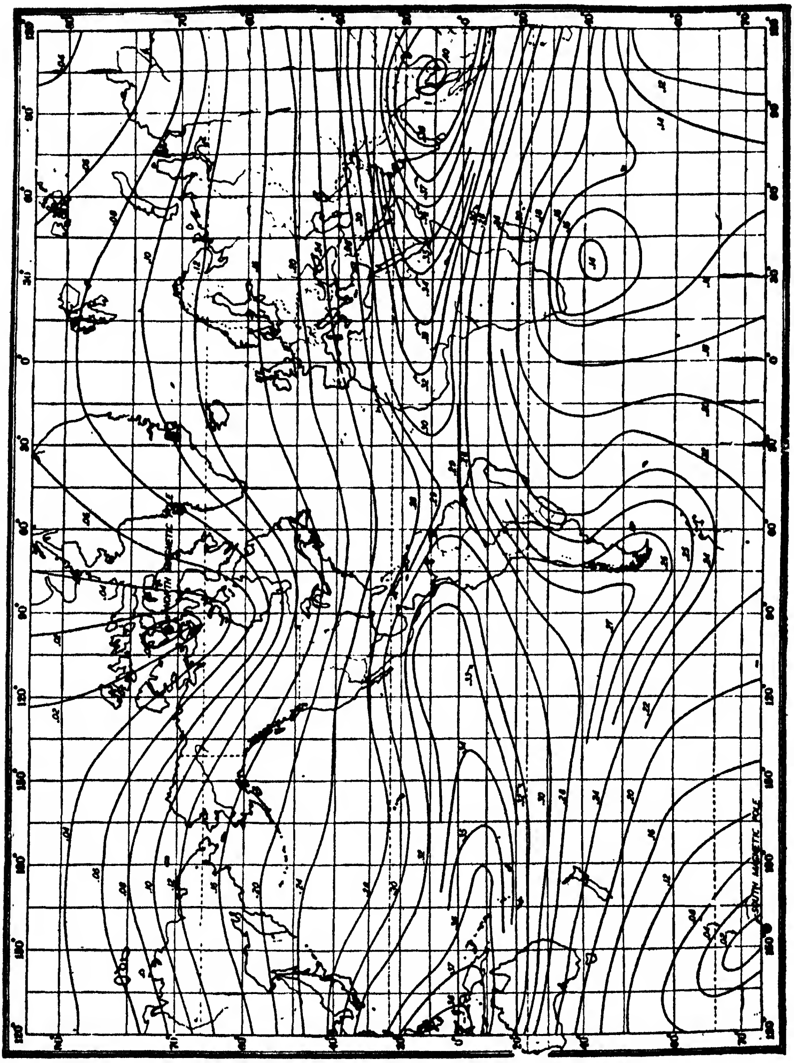


FIG. 31.—Lines of equal magnetic horizontal intensity (isodynamics). (*Courtesy of Carnegie Institution of Washington.*)

considerable amount of material that is magnetizable, there is a possibility that the rotation of the earth on its axis may give rise to the magnetism exhibited by the earth as a whole.

It is also possible that the cause of the earth's magnetism has its origin

quite **outside** the earth itself. In any event the variations in the earth's field are in some way intimately related with solar activity, as already implied. There is reason to believe that the sun emits electrons. Some of these probably reach our atmosphere and thus ionize (Sec. 203) the gases constituting the atmosphere. Such a process would probably give rise to electric currents in the region immediately surrounding the earth, which in turn (Sec. 111) would cause magnetic disturbances. We know that magnetic disturbances always accompany auroral displays, and a study of the spectrum of the aurora would appear to indicate that they are caused by the passage of an electrical discharge through a rarefied gas.

The whole subject of terrestrial magnetism is a most fascinating and important field, made more so because magnetic history is still in the making. The student who is interested in this subject will find a fund of valuable material in papers which appear on this subject in the *Journal of Terrestrial Magnetism and Atmospheric Electricity.*

CHAPTER V

POTENTIAL

41. The Nature of Electrostatic Potential. Potential, in general, is a highly important concept. There are a number of types of potential used in the realm of physics. For instance, we have, among others, gravitational (Newtonian) potential, thermodynamic potential, magnetic potential, and electrostatic potential. What does the term "potential" signify? In general, it may be stated that potential is a quantity the magnitude of which involves energy as a function of **position** or of **condition.** If one can compute the work done in bringing **one unit** of an entity such as mass, heat, or charge to a given point in the region of a quantity of the entity which is being transferred, we shall have a measure of the potential at that particular point. In computing the work, it is necessary to assume some starting point or plane. Sometimes it is convenient to take infinity as the starting point; sometimes the surface of the earth, or a particular condition of the earth, is taken as the point of departure. The meaning of the above statements may be made clear through the use of a simple illustration.

Suppose we have a pebble weighing 1 gm resting on the beach at the **sea level.** In order to lift the stone and place it on a support 1 m above the water would require the expenditure of 980×100 or 9.8×10^4 ergs of energy. In the new position the pebble would have a gravitational **potential** of 9.8×10^4 units with respect to sea level as a plane of reference; and the unit of such gravitational potential would obviously be **ergs per gram.**

At this point it is important to call attention to the fact that **potential and potential energy are not synonymous.** If the stone had weighed 100 gm instead of one, the potential energy in the new position would have been 9.8×10^6 **ergs,** but the potential would still have been only 9.8×10^4 ergs/gm.

Speaking in general terms, it may be said that potential is expressed in terms of the work required to put some unit quantity into a definite position or state. In all cases the force involved is characteristic of that particular system, and the magnitude of the potential is independent of the path over which the unit quantity of the entity is moved. Since potential depends upon position in space, and since it is expressed in terms of energy, it is a scalar quantity.

Turning now to electrical potential we have a situation which is

essentially the same as the gravitational case cited above. In Sec. 7 we indicated that the **condition** of an electrostatic field may be specified in terms of the force acting on unit test charge, and it was also shown that the **status** of the field may be graphically represented by means of lines of force. We are now to consider another method of describing such a field—a method which involves the concept of potential.

Consider, for example, a concentrated positive charge, as shown in Fig. 32. If we attempt to move a test charge between any two points in the field, an expenditure of energy will be involved. Since work is the product of force and displacement a definite amount of work would be done in moving unit positive test charge from p' to p against the repelling force due to the charge q. If the point p' were at infinity the work

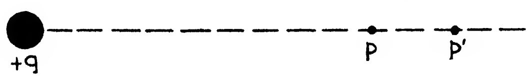


Fig. 32.—A charge gives rise to electrical potential at all points in its field.

involved would be numerically equal to the **electrostatic potential** at p. If the unit test charge were moved only from p' to p, the work done would give the difference of electrostatic potential between these two points. In either case the potential factor would be expressed in **ergs per unit charge.** It follows then that the difference of potential between two points will be unity when 1 erg of work is done in transferring unit charge from one point to the other. If the test charge is moved from a point of lower potential to one at a higher potential, the potential energy of the electrical system will be **increased** and the work will be positive; if moved in the reverse direction the potential energy of the system will be lessened and the work will be considered to be negative.

It was indicated above that the cgs electrostatic unit of potential, and potential difference, is the erg per unit charge. This unit is frequently expressed as ergs per statcoulomb and is designated as a **statvolt.** The volt is the engineering unit of potential; it is equivalent to $\frac{1}{300}$ statvolt. The reason for this particular ratio will be explained later.

From the foregoing it follows that if a quantity of q units of electricity is moved from a point where the potential is V_1 statvolts to a point where the potential is V_2, the work done, in ergs, will be given by the relation

$$W = q(V_2 - V_1). \tag{53}$$

If 1 coulomb of electricity (3×10^9 statcoulombs) is moved between two points whose potential difference is 1 volt ($\frac{1}{300}$ statvolt) the work done, as given by Eq. (53), would be

$$W = 3 \times 10^9 \times \frac{1}{300} = 10^7 \text{ ergs.}$$

Since 10^7 ergs $=$ 1 joule,

$$W \text{ (joules)} = q \text{ (coulombs)} \times \text{p.d. (volts)} \tag{54}$$

where p.d. signifies "potential difference."

42. Potential at Any Point in an Electrostatic Field. It will be found useful to have available an expression by the use of which one may determine the value of the potential at any point in an electrostatic field.

Suppose we have a concentrated charge as shown in Fig. 33. Con-

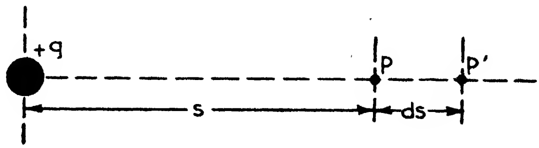


FIG. 33.—Electrical potential at any point in a field, expressed in terms of work.

sider a point p' located a very short distance ds from p. Since ds is very small, the intensity will be practically constant between p and p' and will be given by q/Ks^2. It therefore follows that the work necessary to transfer unit test charge from p' to p will be given by the relation

$$dw = -\mathcal{E}\,ds = -\frac{q}{Ks^2}\,ds.$$

The negative sign indicates that the displacement and the field intensity are oppositely directed. If now we bring our test charge from infinity to p, along a radius vector, the work done in that operation will be

$$\int_\infty^s dw = -\frac{q}{K}\int_\infty^s \frac{ds}{s^2}.$$

This, by definition, is numerically equal to the potential at p; hence we may write

$$V_p = -\frac{q}{K}\int_\infty^s \frac{ds}{s^2} = \frac{q}{Ks}. \tag{55}$$

In the above relation, if q is expressed in statcoulombs and s in centimeters, the potential V will be given in statvolts.

In most cases, one is interested in the potential **difference** between some two points, both of which are at finite distances. In that event, the equation would take the form

$$V_p - V_{p'} = -\frac{q}{K}\int_{s_2}^{s_1} \frac{ds}{s^2} = \frac{q}{K}\left(\frac{1}{s_1} - \frac{1}{s_2}\right) \tag{56}$$

where s_1 and s_2 represent the distance shown in Fig. 34. Here again, if q is in statcoulombs, the potential difference will be in statvolts.

In any case where the direction of the displacement is not along the line of the radius vector, one can, by referring to Fig. 35, write an expression for potential as follows,

$$V_p = \int_{\infty}^{s} \frac{q}{Ks^2}\, dl \cos \alpha.$$

But $dl \cos \alpha = ds$, hence the above equation reduces to the form from which Eq. (55) was derived. Here, then, we have analytical proof that

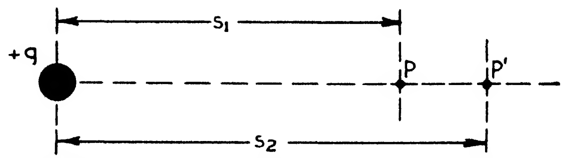


FIG. 34.—Difference of potential between any two points in a field.

the potential at a point is a **space property** of that point, and its **magnitude is independent** of the direction of the path along which the unit test charge is transferred.

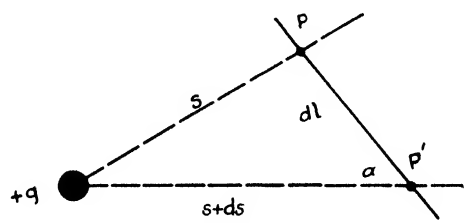


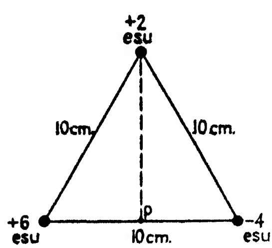
FIG. 35.—Illustrating the fact that potential is a space property of the point in question and not a function of direction.

If the total charge is not concentrated but is distributed in discrete quantities located at various centers, we may find the total potential by taking the scalar sum of the potentials due to each component charge. Such a case could be expressed thus,

$$V_p = \frac{q_1}{Ks_1} + \frac{q_2}{Ks_2} + \frac{q_3}{Ks_3} + \cdots$$

$$= \sum \frac{q}{Ks} \quad \text{ergs/unit charge, or statvolts.} \qquad (57)$$

Problem. Three charges are located in air as shown in the accompanying sketch. Find the potential at the point p midway between the $+6$ and -4 charges.



Solution.

$$V_p = \frac{6}{5} + \frac{2}{8.66} + \left(-\frac{4}{5}\right)$$
$$= 0.633 \text{ statvolt}$$
$$= 189.9 \text{ volts.}$$

43. Potential Gradient. A simple but widely used relation exists between potential and field intensity. In the development of Eq. (55) we made use of the expression

$$dw = -\mathcal{E}\,ds.$$

By definition this may also be written

$$dv = -\mathcal{E}\,ds.$$

Changing the form, we have

$$\mathcal{E} = -\frac{dv}{ds}. \tag{58}$$

In words, Eq. (58) means that the **intensity at any point in an electrostatic field is numerically equal to the negative space rate of change of potential at that point.** The quantity dv/ds is frequently referred to as the **potential gradient** of the field. If the potential difference dv is expressed in statvolts and ds in centimeters, the potential gradient will be in statvolts per centimeter; if dv is in volts and ds in centimeters the gradient will be in volts per centimeter. The physical meaning of the negative sign is that as s increases, V decreases.

An equivalent form of the above relation may be derived in another way. If one moves unit test charge between two points which are at different potentials and separated by a distance s, it would follow from definition that

$$V_2 - V_1 = \int_{x=0}^{x=s} F\,dx.$$

Hence

$$V_2 - V_1 = Fs.$$

Since we are here dealing with unit charge it follows that the field intensity is **numerically** equal to the mechanical force involved, *i.e.*, $\mathcal{E} = F$. Combining these two expressions, we have

$$\mathcal{E}s = V_2 - V_1.$$

This may be written

$$\mathcal{E} = \frac{V_2 - V_1}{s}, \tag{59}$$

which is an equation giving the potential gradient in a convenient and

useful form. Thus it is seen that we can arrive at the value of the field intensity by either of two methods: (1) by finding the force on unit test charge, or (2) by determining the fall in potential per centimeter.

As an example of the use of this relation, it may be stated that the normal potential gradient of the atmosphere over land varies from 67 to 317 volts/m; over the sea the gradient is of the order of 128 volts/m.

44. Magnetic Potential. Magnetic potential at a point is defined as the work done in moving unit N pole from infinity to the point in question; and difference of magnetic potential is the work done in moving unit N pole between the two points involved. The unit of magnetic potential and potential difference is the **erg per unit pole.**

Following the reasoning employed in dealing with electrostatic potential (Sec. 42), it may be said that the magnetic potential at any given point in a field due to the single pole will be given by the relation

$$V = -\int_{\infty}^{s} \frac{m}{\mu s^2}\, ds = \frac{m}{\mu s} \qquad \text{ergs/unit pole.} \tag{60}$$

There is no special name for the unit of magnetic potential.

If the field is due to the presence of several poles rather than one, the potential at a point would be

$$V = \frac{m_1}{\mu s_1} + \frac{m_2}{\mu s_2} + \frac{m_3}{\mu s_3} + \cdots = \sum \frac{m}{\mu s}. \tag{61}$$

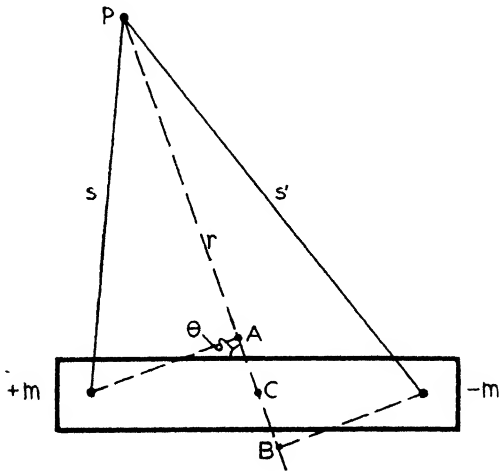


FIG. 36.—Magnetic potential at a point, due to a dipole.

The value of the potential at some point in the field of a magnetic dipole, as indicated in Fig. 36, is sometimes desired. By Eq. (61) the

potential at p would be

$$V_p = \frac{m}{s} + \left(-\frac{m}{s'}\right).$$

It is possible to express the pole distances in terms of the length of the magnet and the angle θ made by a line from p to the center of the magnet. Let $r = PC$ and $x = l/2$, half the length of the magnet. By erecting perpendiculars on pB we have that $AC = CB = x \cos \theta$. Substituting in the above equation for V_p, and assuming that l is small compared with r, we find that

$$V_p = \frac{m}{r - x\cos\theta} - \frac{m}{r + x\cos\theta} = \frac{2mx\cos\theta}{r^2 - x^2\cos^2\theta}$$
$$= \frac{ml\cos\theta}{r^2 - x^2\cos^2\theta} = \frac{M\cos\theta}{r^2 - x^2\cos^2\theta}.$$

At any appreciable distance from a dipole the interpole distance $l(=2x)$ will be negligible in comparison with r; hence, to a first approximation, the last equation above becomes

$$V_p = \frac{M}{r^2}\cos\theta, \tag{62}$$

where M is the magnetic moment of the dipole.

As was noted in dealing with an electrostatic field, the field intensity in the magnetic case may be found by differentiating the potential with respect to the distance factor. The intensity along the radius vector would accordingly be

$$H_r = -\frac{\partial V}{\partial r} = \frac{2M}{r^3}\cos\theta. \tag{63}$$

The partial differential expression is used above because all factors other than r are held constant.

For points on the axis of the magnet $\theta = 0$; hence

$$H_r = \frac{2M}{r^3}, \tag{64}$$

which is identical with the result obtained for the Gaussian A position in Sec. 26, Eq. (26).

From the foregoing discussion it is evident that **magnetic field intensity is numerically equal to the maximum space rate of the change in potential in any specified direction.**

45. Equipotential Lines and Surfaces. In Sec. 42 [Eq. (55)] we found that the electrostatic potential at any point in a field varies inversely as the distance from the charge giving rise to the field. It follows there-

fore that, in the case of an isolated concentrated charge, all points at a certain distance from the charge will be at the same potential. Using the distance from the charge to some definite point as a radius, one may describe a surface about the charge and all points on this spherical surface will be at the same potential. Such a surface is known as an equipotential surface; it is the locus of all points having the same potential. If we are dealing with one dimension only, a line connecting all points having the same potential would be what is called an **equipotential line.**

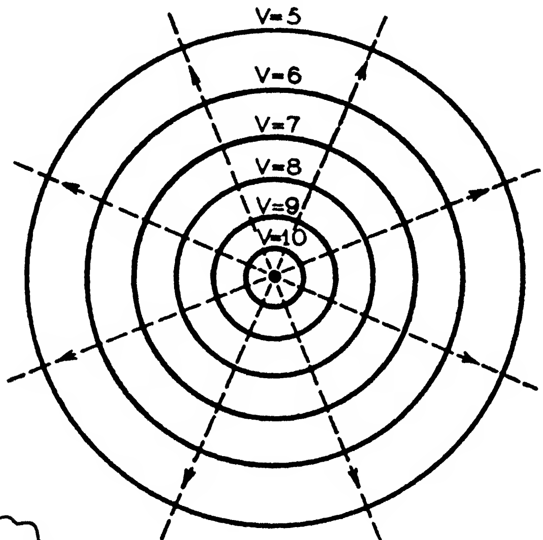


FIG. 37.—Equipotential lines about a concentrated charge. The spacing represents a potential difference of 1 volt.

Figure 37 shows equipotential lines drawn about a concentrated charge, the distance between concentric circles representing a potential difference of one unit. (Why does the distance between the successive equipotentials gradually increase?)

In general, charges will not be concentrated but will be distributed nonuniformly through space or over a surface. In such cases the equipotential surfaces surrounding the collection of charges will not be spheres, and the equipotential lines will not be circles. Figure 38 shows the equipotential line pattern in the case of two neighboring unlike charges. The dotted lines indicate equipotentials, and the continuous lines cutting the equipotential lines are the lines of force.

The following points should be noted: (1) No work is done in moving a test charge from one point to another point on an equipotential surface, but energy must be expended to move a charge from one surface to

another. (2) **Lines of force are at all points normal to equipotential surfaces.** This follows from the fact, stated above, that no work is involved in displacing a test charge along an equipotential surface. Therefore, the direction of motion along such a surface must be at right angles to the lines of force. (3) No two lines (equipotential) or surfaces can intersect; if they did the potential function would not be single-valued. If

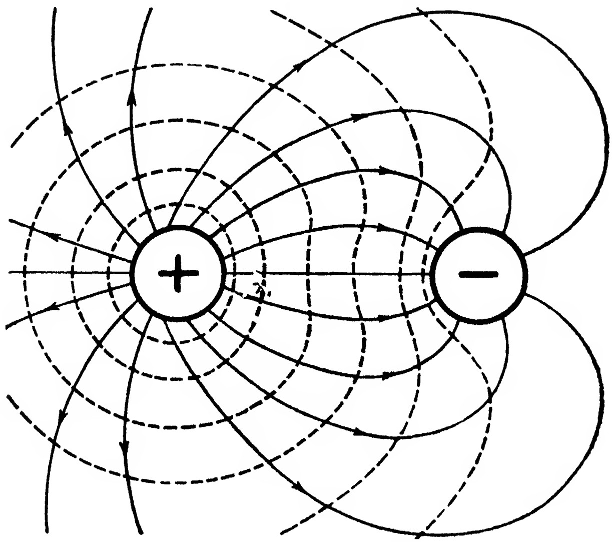


FIG. 38.—Equipotential lines (*dotted*) and lines of force in the region of two unlike charges.

the charges on a conductor are in equilibrium the surface is an equipotential surface. (4) All points within a conducting shell or screen are at the same potential. We have seen (Sec. 12) that the field intensity at all points in the region surrounded by a conducting screen is zero. Under these circumstances Eq. (58) takes the form

$$\frac{dv}{ds} = 0,$$

and integration of this expression gives a constant. Therefore, a region within a closed conducting surface is an equipotential region.

In the foregoing discussion we have been considering electrostatic fields but the principles enunciated are equally applicable to the magnetic case. We may therefore speak of magnetic equipotential lines and surfaces in a similar manner.

The relation between equipotential surfaces and lines of force is of great practical importance, notably in the field of electron optics.

46. Potential Energy of a Charge. In charging a body—*i.e.*, in causing a redistribution or rearrangement of the electrons—energy must be expended. The charge represents, therefore, a definite amount of potential energy, and it is frequently convenient to be able to compute the amount of this energy. If we can determine the amount of work done in placing the charge in its position we will have a measure of the potential energy represented by the charge. The method of analytical attack is similar to that followed in determining the magnitude of the potential energy of an elevated quantity of water.

Let us suppose that we bring a number of very small elemental charges dq from some point that is at zero potential and place them one at a time upon an insulated conductor, the initial potential of which is also zero. As the charge on the body increases its potential will rise [Eq. (55)]; therefore the amount of work done in transferring each succeeding increment will be greater than for the one preceding. In Sec. 42 it was pointed out that the potential due to a charge is measured by the work required to transfer unit charge from a point at zero potential to the point in question, and it has also been shown [Eq. (53)] that the work necessary to move any charge will be the product of the potential difference and the charge. In this case, however, the potential increases from zero to some value V. But the average potential would be $V/2$. The work done in placing a total charge Q on the body would therefore be given by

$$W = \frac{V}{2}\int_0^Q dq = \tfrac{1}{2}VQ. \tag{65}$$

Since this represents the work done in charging the body, it must also represent the potential energy of the charge. If V is in statvolts and Q in statcoulombs, W will be in ergs. By introducing the necessary transformation factors we may express the potential in volts and the charge in coulombs. Doing this, the right side of Eq. (65) would take the form

$$\tfrac{1}{2}[V\text{ (statvolts)} \times 300]\,\frac{Q\text{ (statcoulombs)}}{3 \times 10^9} = \tfrac{1}{2}\,\frac{VQ}{10^7}.$$

In order to preserve the equality in our original equation it will be necessary to divide the left side also by 10^7, which in turn will reduce the ergs to joules. Hence we may write

Potential energy represented by a charge

$$= W\text{ (joules)} = \tfrac{1}{2}V\text{ (volts)} \times Q\text{ (coulombs)}. \tag{66}$$

Problem. A conducting body bears a charge of 0.02 coulomb. It is found that the body has a potential of 1,000 volts. What potential energy does the charge represent?

Solution. $W = (1{,}000 \times 0.02)/2 = 10$ joules.

In concluding our discussion of potential it is well to emphasize the distinction between **electrical potential** and **electrical potential energy.** The former is measured by the work involved in bringing **unit test charge** from infinity (or from a point where the potential is zero) to the charged body; while the latter has to do with the **total work done in giving the body its complete charge.** It is highly important that this significant distinction be kept clearly in mind as we proceed.

PROBLEMS

1. What is the potential, in volts, at a point in air 25 cm from a concentrated charge of 5 statcoulombs? What would be the potential at the same point if the surrounding medium was oil ($K = 2.5$)?

2. How much work would be done in moving a charge of 2×10^{-6} coulomb from a point where the potential is 0.5 statvolt to one where the potential is 5 statvolts?

3. Three concentrated charges of +4, +6, and −10 statcoulombs, respectively, are so located that lines joining them form an equilateral triangle whose sides are 20 cm. Compute the potential at the center of the line joining the 4- and 6-unit charges, assuming air to be the dielectric.

4. Compare the magnitude of the potential in the region surrounded by a conducting shell with the value of the potential on the outside surface of the shell.

5. It is found that the potential difference between two points is 100 volts. One of the points is 25 cm from the charge involved. What is the magnitude of the charge which is giving rise to the field?

6. Two points 30 m apart are at different levels in the atmosphere and show a potential difference of 10 statvolts. What is the potential gradient in volts per meter?

7. If the potential difference between a cloud and the earth is 10^9 volts, and if it be assumed that the quantity of electricity constituting a lightning flash is 25 coulombs, how many foot-pounds of work could be done if all of the energy represented by the discharge could be utilized? If the flash occupied 1 sec of time, what horsepower will the discharge represent?

8. What is the magnetic potential at a point 100 cm from the + pole of a magnet on a line perpendicular to the axis of the magnet and passing through that pole? Assume the poles to be of equal strength, and having a value of 5 units.

9. Using the general expression for magnetic potential given by Eq. (60), show that one may derive an expression for the magnetic field intensity in the Gaussian B position. The result should correspond to Eq. (28).

10. Suppose a wire is surrounded by a conducting cylinder, the outside of which is at ground potential (zero). Assume that the space between the wire and the sheath is filled with a compound whose dielectric constant is K. Further, let us suppose that the wire carries a charge of q statcoulombs per centimeter of length. Take the diameter of the wire to be r and the inside radius of the inclos-

ing cylinder to be r'. Derive an expression for the difference of potential between the core and the sheath.

11. Having derived a working equation, as suggested in Prob. 10, apply it to the following specific case: $K = 4$; $r = 1$ mm; $r' = 5$ mm; $q = 10$ statcoulombs. Find the potential difference between the core and the sheath.

12. Three equally spaced positive point charges are located on the surface of a nonconducting sphere. Neglecting any effect due to the insulating material, what will be the potential at the center of the sphere? The charges are 100 statcoulombs each, and the sphere is 20 cm in diameter. What will be the potential at a point 50 cm from the center of the sphere?

13. Point charges of +20, −40, +60, and −80 statcoulombs are located at the corners of a 12×12 cm rectangle. What is the potential at the center of the rectangle?

14. Under the conditions specified in the last problem, locate a point where the potential is zero.

15 A charged oil drop whose mass is 2×10^{13} gm is held in a fixed position between two charged parallel plates positioned 0.5 cm apart in air. Assuming that the charge carried by the drop consists of 2 electrons, what potential difference must be maintained between the plates in order that the drop shall remain at rest?

16. It is found that 25 joules of work is done in charging a body to 10,000 volts. What is the magnitude of the charge involved?

17. What is the magnetic field strength at a point 3 cm from the + pole of a magnet which is 5 cm long, the point being at right angles to the major axis of the magnet? Assume that the pole strength (both poles) is 25 units.

18. What will be the magnetic potential at the point specified in the last problem?

CHAPTER VI

MEASUREMENT OF POTENTIAL

47. Absolute Electrometer. The determination of potential magnitudes in connection with various research and engineering problems constitutes one of the most important electrical measurements which one is called upon to make. It is not experimentally feasible to measure potential or potential differences directly, *i.e.*, to measure the work done in moving unit test charge from one point to another. Therefore, indirect methods must be employed. One procedure followed makes use of the

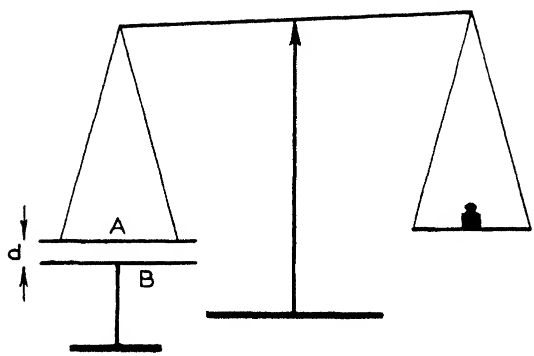


FIG. 39.—Showing the principles involved in the operation of the attracted-disk electrometer.

fact that two oppositely charged plates will attract one another. It is possible to find a relation between this force of attraction and the potential difference existing between the two plates.

A simple device utilizing this principle is sketched in Fig. 39. This organization, devised by Sir William Harris, is known as an **attracted-disk electrometer.** If, by any convenient means, a difference of potential is established between the two plates, the resulting electrostatic attraction will cause the movable plate to move toward the corresponding fixed member. We can compensate for this displacement by adding weights to the right pan of such a specially constructed balance. When mechanical equilibrium has been restored the weight added to the pan will exert a gravitational force mg equal to the electrostatic force existing between the plates. We are then in a position to find a working relation by means of which one may evaluate a potential difference **in terms of the three fundamental quantities, length, mass, and time.**

In Sec. 15 we found that the mechanical **force per unit area** exerted by a charged plate on a second parallel and oppositely charged plate is given by the expression [Eq. (13)]

$$F = \frac{\mathcal{E}^2 K}{8\pi},$$

where $\mathcal{E}$ is the field intensity and K the dielectric constant. In this case $K = 1$, and for some definite area A the equation would become

$$F = \frac{\mathcal{E}^2 A}{8\pi}.$$

In this case Eq. (59) will take the form

$$\mathcal{E} = \frac{V_A - V_B}{d},$$

where $V_A - V_B$ represents the potential difference between the plates and d the distance of separation. Eliminating $\mathcal{E}$ between the last two equations we have

$$F = mg = \frac{A(V_A - V_B)^2}{8\pi d^2} \quad \text{dynes} \tag{67}$$

as the mechanical force acting on the movable plate A. In practice this force is found in grams and converted into dynes. Solving the above equation for potential difference we get

$$V_A - V_B = d\sqrt{\frac{8\pi F}{A}} = d\sqrt{\frac{8\pi mg}{A}} \tag{68}$$

as a working relation. If d is in centimeters and A in square centimeters, the potential will be in statvolts.

Problem. It is found that 0.02 gm is required to restore equilibrium when the plates of an electrometer are connected to a source of potential difference. The distance between the plates is 0.5 cm and the effective area of the movable plate is 20 cm². If g is taken as 980 cm/sec², what is the potential difference between the plates?

Solution.

$$\text{p.d.} = 0.5\sqrt{\frac{8\pi 0.02 \times 980}{20}} = 0.0248 \text{ statvolts} = 7.44 \text{ volts}$$

In the case of a plate, such as that illustrated, the electrostatic field will not be uniform over the entire surface; the lines of flux will spread outward near the edges. As a result of this the simple form of the electrometer just described will not give accurate results. This defect is corrected in the instrument devised by Lord Kelvin. The essentials of

the Kelvin absolute electrometer are shown in Fig. 40. In the Kelvin instrument a "guard ring" *G* has been added which is electrically connected to the movable disk, and which is separated from the disk by the smallest possible space. As a result of the position of this guard ring, the lines of electric flux are normal to the central disk, even near the edges, and hence the field is uniform between the plates. Delicate and accurate means are provided for determining the exact position of equilibrium.

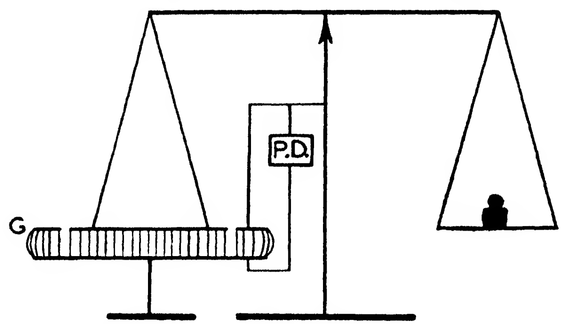


Fig. 40.—Diagrammatic sketch of the essentials of the Kelvin electrometer.

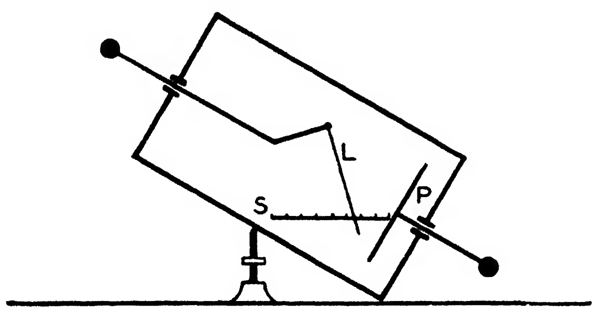


Fig. 41.—Essentials of the adjustable-leaf electroscope.

The electrometer described above is an **absolute instrument** in that the determination of the potential difference depends **only** upon the measurement of certain fundamental quantities and **not** upon the magnitude of another like quantity. This is an important distinction. Because the Kelvin electrometer is an absolute instrument it can be, and is, used as the ultimate reference standard in connection with the calibration of other potential-indicating instruments.

48. The Electroscope. There are a number of other electrometers in common use, but they all give only **relative** value of potential, *i.e.*, they must first be calibrated by reference to known potential values.

One such device is the metallic-leaf electroscope encountered in the elementary study of electricity. A more sensitive form of this instru-

ment is sketched in Fig. 41. This instrument does not differ essentially from the ordinary gold-leaf unit. However, by providing means whereby the whole instrument may be tilted, thus varying the resting distance between the leaf L and the plate P, the unit becomes extremely sensitive. It must of course be first calibrated by applying known potential differences to the system. A practical form of this type of electrometer is shown in Fig. 42.

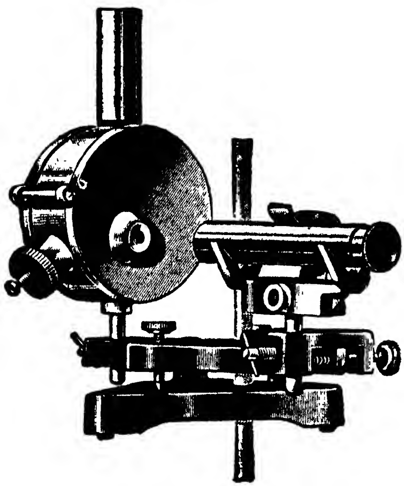


FIG. 42.—Practical form of the tilting electroscope. (*Central Scientific Co.*)

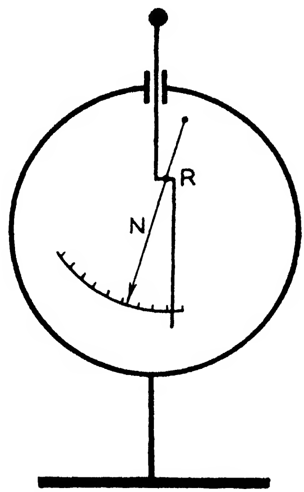


FIG. 43.—Essentials of the Braun electrometer. The pivoted metallic needle is repelled by the fixed member with which it makes electrical contact.

49. Braun Electrometer. Another type of comparison electrometer, due to Braun, is in reality a modified form of the metallic-leaf form of instrument. As seen in Fig. 43, the electrical system consists of a light flat needle which is pivoted on a conducting support. As in the simple gold-leaf electroscope, any potential applied to the external connection will cause the support R and the needle N to acquire like charges. As a result, the needle will be repelled and will assume a position of electromechanical equilibrium depending on the magnitude of the applied potential. The Braun instrument is calibrated, by reference to known potential values, to read directly in volts. It is useful in dealing with potentials of the order of several thousand volts. A commercial form of this type of electrometer, called the Braun electrostatic voltmeter, is shown in Fig. 44.

50. Quadrant Electrometer. An instrument which has been widely used in scientific laboratories for measuring potential **by the comparison method,** and which has served as the basis for a portable engineering instrument, is one which is also due to Lord Kelvin. The quadrant electrometer takes its name from one of the mechanical features of its construction. The instrument, in its modern form, consists essentially of a specially shaped very light vane, or "needle," supported by a metallized quartz fiber, the vane being suspended within a metal box. This box

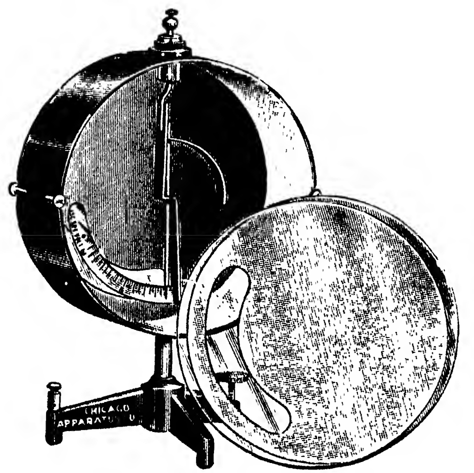


FIG. 44.—Practical form of the Braun electrometer. The deflection of the movable aluminum vane is a function of the potential of the fixed member. The instrument is calibrated to read directly in volts. (*Chicago Apparatus Co.*)

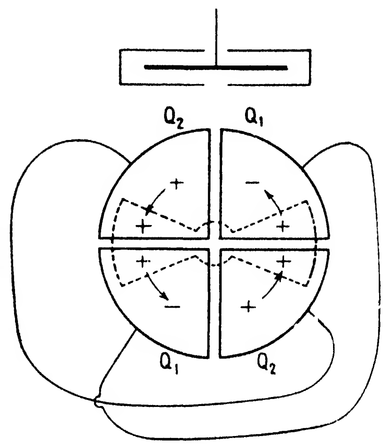


FIG. 45.—Diagrammatic top view of the essential components of the quadrant electrometer. The suspended vane is shown dotted.

is divided into four parts or "quadrants," the opposite quadrants being electrically connected. Figure 45 indicates the movable and fixed elements of one form of this instrument. The quadrants are supported by insulating pillars made of quartz or amber. The movable vane is of very thin aluminum or paper coated with thin foil. A mirror is attached to a light wire which is fastened to the vane. This particular form of the quadrant electrometer is due to F. Dolezalek.

In use the vane or needle is charged by being connected to a source of potential as, for example, the positive side of a 100-volt storage battery, the negative battery terminal being grounded. The source of the difference of potential that is to be measured is connected to opposite pairs of quadrants. With no charge on the quadrants, the vane is adjusted to hang symmetrically with respect to the quadrants, as shown in Fig.

45. With the quadrants oppositely charged, the needle will be repelled by one set of quadrants and attracted by the other, as indicated by the arrows in Fig. 45. The movement continues until the electrostatic torque is balanced by the mechanical torque of the suspension. The deflection, as shown by a telescope and scale, is compared with the deflection due to a known difference of potential on the quadrants. With a suitable suspension, a difference of potential of 1 volt on the quadrants may be

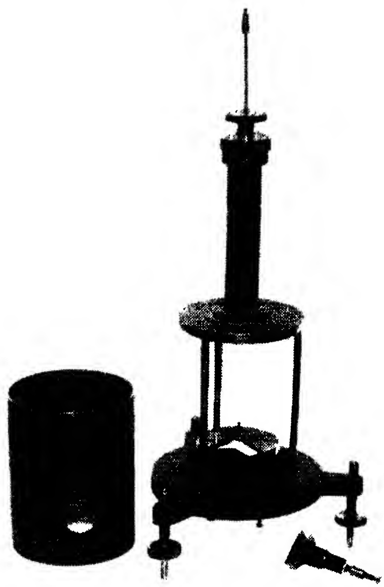


FIG. 46.—Modern form of the quadrant electrometer. One quadrant has been removed to show suspended vane.

made to give a deflection of 200 to 400 mm on a scale placed 1 m from the mirror. Figure 46 shows a late model of this type of electrometer.

Because of its wide utility in research, and also because the theory of the electrometer so well illustrates the method of analytical attack in problems in electrostatics, it is worth while to outline the development of the working formula of the instrument.

In the diagrammatic sketch shown as Fig. 45, suppose the electrical conditions to be as indicated. The needle will then tend to move as shown by the arrows. Equilibrium will obtain when the mechanical couple equals the couple due to the electrostatic field. Our problem is to find the relation between the deflection and the potentials of the essential components of the instrument.

To do this let us suppose that the vane has been slightly displaced due

to the electrostatic force action. For small deflections the restoring or mechanical couple L is proportional to the deflection ϕ. Since the moving system has been displaced from its original neutral position, the total energy of the charged system must have been changed. If the magnitude of this change can be found, we can develop an expression for the work done in effecting the displacement, which in turn will be proportional to the displacement.

We have shown [Eq. (65)] that the potential energy of any electrified system is given by $\frac{1}{2}QV$, and also [Eq. (8)] that $Q = \sigma A$ where A is the area involved and σ the surface density of the charge. Combining these two relations we have

$$\text{P.E.} = \tfrac{1}{2}\sigma A V. \tag{i}$$

By comparing the values for field intensity as given by Eqs. (11) and (59), we find that

$$\sigma = \frac{V}{4\pi d}, \tag{ii}$$

where V is the potential difference between the two plates of a system and d their distance apart.

Substituting (ii) in (i) we get, as a general expression for the potential energy of the system,

$$\tfrac{1}{2}\,\frac{V^2A}{4\pi d}.$$

In our case, we have changed the effective area by an amount da; hence the change in energy (**increase** in this case) would be

$$\frac{V^2}{8\pi d}\,da.$$

In the electrical system being considered, V is $V_N - V_{Q1}$; where V_N is the potential of the vane and V_{Q1} the potential of one pair of quadrants. Thus the increase in energy is

$$\frac{da}{8\pi d}\,(V_N - V_{Q1})^2.$$

If desired, da may be expressed in terms of the actual dimensions of the vane, but since these, together with d, are constants of the particular instrument being used, we may leave this factor in its present form.

Similarly, as the vane moves slightly **away** from the negative quadrants the system suffers a **decrease in energy**. This decrease will, by analogy, be equal to

$$\frac{da}{8\pi d}\,(V_N - V_{Q2})^2.$$

The total energy change will be the difference between the last two expressions, or

$$\frac{da}{8\pi d}[(V_N - V_{Q1})^2 - (V_N - V_{Q2})^2],$$

which reduces to

$$\frac{da}{4\pi d}(V_{Q2} - V_{Q1})\left(V_N - \frac{V_{Q1} + V_{Q2}}{2}\right).$$

This should equal the work done in bringing about the change in position of the vane.

Now for small displacements the work done will be proportional to the deflection ϕ, or $W = B\phi$, where B is a constant. Therefore

$$B\phi = \frac{da}{4\pi d}(V_{Q2} - V_{Q1})\left(V_N - \frac{V_{Q1} + V_{Q2}}{2}\right);$$

hence

$$\phi = \frac{da}{4\pi dB}(V_{Q2} - V_{Q1})\left(V_N - \frac{V_{Q1} + V_{Q2}}{2}\right).$$

All of the factors in the term $da/4\pi dB$ are constants, and depend upon the construction of the particular instrument in use. In practice V_N is large compared with the potential of the quadrants; hence the second bracketed term does not differ materially from V_N. It therefore follows that the deflection ϕ is proportional to

$$(V_{Q2} - V_{Q1})V_N, \tag{69}$$

i.e., to the product of the potential of the vane and the potential difference between the quadrants. We thus have another, and an accurate, method of **comparing differences of potential.**

It frequently becomes necessary to determine alternating differences of potential. By connecting one pair of quadrants with the vane the deflection will be in the same direction, regardless of whether either component of the system is positive or negative. When connected in this manner, the electrometer may be utilized in a-c work. When so arranged the working equation reduces to a form which shows that the deflection is proportional to

$$\tfrac{1}{2}(V_N - V_{Q1})^2 \tag{70}$$

where V_N is the potential of the pair of quadrants that is connected to the vane. In general, this arrangement is less sensitive than the first form of connection. When utilized in this manner, the instrument is said to be used **idiostatically.** When the quadrants and the vanes are all at different potentials, the instruments is said to be operating **heterostatically.**

51. Electrostatic Voltmeters. By utilizing the principle of the quadrant electrometer when used idiostatically, it is possible to devise a portable voltmeter suitable for both direct and alternating potentials. Lord Kelvin designed such an instrument, one form of which is shown in Fig. 47. In this particular type of the instrument control is effected by means of an adjustable weight attached to the movable vane or by a delicate helical spring attached to the shaft which supports the moving

FIG. 47.—Electrostatic voltmeter, low-range type.

system. The scale is calibrated directly in volts, the range in some cases extending up to several thousand volts.

For measuring potentials of the order of 100 volts Lord Kelvin developed what is known as a multicellular voltmeter. This differs from the above in that there are a number of vanes alternating with corresponding sets of quadrants. The vanes occupy a horizontal position and are attached to a light vertical spindle that, in turn, is suspended by means of a short conducting wire or ribbon. The deflection is observed by means of a light pointer attached to the movable system and moving over a horizontal scale. The chief advantage of the electrostatic voltmeter is that, once the moving system comes to rest, no energy is taken from the source. When used in making observations, a slight energy consumption

is involved in alternating-potential measurements. Commercial electrostatic voltmeters are available which will read to 200,000 volts.

52. Compton Electrometer. Probably the most sensitive form of the quadrant electrometer that has been devised is that developed by Dr. Karl T. Compton.[1] A high degree of sensitivity is secured by giving the needle a slight tilt about its long axis, and also by vertically displacing one pair of quadrants with respect to the other. Figure 48 is a diagrammatic side view of the Compton system. As a result of this asymmetrical arrangement of the needle and the quadrants, electrostatic forces are brought into play which may be made to neutralize, more or less completely, the mechanical torque of the suspension. It is possible with this instrument to secure a sensitivity as high as 50,000 mm/volt.[2] However, the ordinary working range is 0 to 10,000 mm/volt. Deflections are proportional to the potential difference over a relatively wide range.

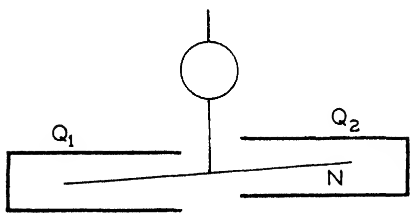


FIG. 48.—Showing essential elements of a Compton electrometer.

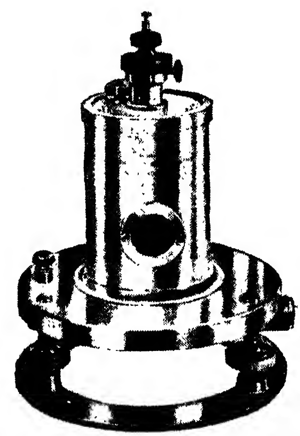


FIG. 49.—Practical form of the Compton electrometer. (*Cambridge Instrument Co.*)

The outstanding features of the Compton form of electrometer are the small size of the electrical system (the needle is about 1 cm in length), the resulting small capacitance (Sec. 55), and the ease with which the sensitivity may be adjusted over a wide range. When adjusted for high sensitivity, neither the uniformity of the scale nor the rapidity with which reading may be taken are unduly sacrificed. Figure 49 shows an

[1] See a paper by the Messrs. Compton, A Sensitive Modification of the Quadrant Electrometer, *Phys. Rev.*, **14** (No. 2), 85 (1919).

[2] The term "sensitivity" as here used refers to the deflection observed by means of a telescope and scale when a potential difference of 1 volt is applied to the quadrants, the scale being at a distance of 1 m from the mirror. A similar meaning is often attached to the term "sensitivity" in other connections.

exterior view of one model of the Compton instrument. Because of its simplicity and high sensitivity, the Compton electrometer has been extensively employed in research work involving accurate electrostatic measurements.

53. String Electrometer. For certain purposes, particularly in research work, the string electrometer is admirably adapted to the accurate measurement of small values of potential. The essential parts of an instrument of this type are shown diagrammatically in Fig. 50. The

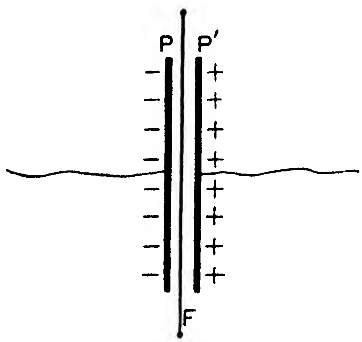


Fig. 50.—Essentials of the string electrometer.

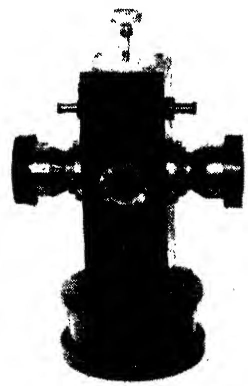


Fig. 51.—Practical form of the string electrometer. The deflection of the movable element is read by means of an attached microscope that carries a scale in the eyepiece.

device consists of two carefully insulated plates P and P' between which is stretched an insulated quartz fiber F which is metallized. If the string is connected to a source of potential, and a difference of potential applied to the plates, the string will be slightly deflected as a result of the electrostatic force action. The deflection is read by means of a micrometer microscope that forms a part of the instrument. The position of the plates and the tension of the string are controlled by means of micrometer screws. Dr. Joel Stebbins has designed an electrometer of this type that possesses a number of desirable features. The Stebbins instrument has an extremely small temperature coefficient, high sensitivity, small electrostatic capacitance, and a short period. An exterior view of this instrument is shown in Fig. 51.

54. Needle and Sphere Spark Gaps. When dealing with potentials of the order of 10 to 200 kv, use is sometimes made of the electrical breakdown characteristics of air as a dielectric. When the potential between two sharp points reaches a high value, a disruptive discharge takes place between the points. There is an approximate relation between the separation of the points and the potential difference at which a discharge begins to take place. The American Institute of Electrical Engineers has laid down specifications as to the physical dimensions of the needle points and the temperature, barometric pressure, and relative humidity under which such a method may be used. The needle gap is used for voltages from 10 to 50 kv.

Above 50 kv, a gap consisting of carefully machined and polished spheres gives fairly reliable results when used under specified conditions. For instance, if 100 kv is applied to a pair of spheres 25 cm in diameter breakdown will occur when the gap length is 5.2 cm; and at 200 kv the rupture of the air dielectric will occur at 12.8 cm. The American Institute of Electrical Engineers specifications governing these methods of measuring high potentials will be found in any standard electrical engineering handbook.

In later chapters other methods of determining potential differences will be described.

CHAPTER VII

CAPACITANCE

55. Concept of Capacitance. We have already shown (Sec. 41) that the potential of a conductor is measured by the work done in bringing unit test charge from infinity up to the conductor. It has also been deduced (Sec. 11) that the field strength, and hence the opposing force, is proportional to the charge on the conductor. It therefore follows that the potential of an insulated conductor is proportional to its charge. This fact may be stated in another way by saying that the ratio of the charge to the potential is constant, thus

$$\frac{Q}{V} = C. \tag{71}$$

If we rewrite our equation in the form

$$V = \frac{Q}{C},$$

it will be evident that the potential of a conductor depends not only upon the charge resident thereon, but also upon a second factor C which we may designate as the **capacitance** of the conductor.

Equation (71) indicates that **we may define capacitance as the ratio of the charges on a conductor to its potential.** The same relation also shows that capacitance is numerically equal to the charge that will cause a conductor to have unit potential. In terms of mechanics, it would be analogous to the number of gas molecules that it would be necessary to introduce into a given container in order to establish unit pressure. It is obvious that the amount of gas required to produce unit pressure would depend, in part, upon the volume of the container. In a somewhat analogous manner, it is determined that the capacitance of a conductor is a function of its superficial area and of its geometrical form, as will be shown in discussions which follow.

Further, since the potential of a body depends not only upon its own charge but also upon the presence of other charged bodies in its vicinity, it follows that the capacitance of a conductor will depend upon its position with respect to other charged bodies. This will be more apparent when we deal with the subject of condensers in a later section.

56. Capacitance of a Sphere. We shall find it convenient to be able to compute the capacitance of conductors, and assemblies of conductors,

having definite geometrical forms and relative positions. We may begin by deducing an expression which will give the capacitance of an electrically isolated spherical conductor surrounded by a dielectric. Let the situation be as sketched in Fig. 52. We will assume that the sphere possesses a charge whose value is Q statcoulombs. The potential V

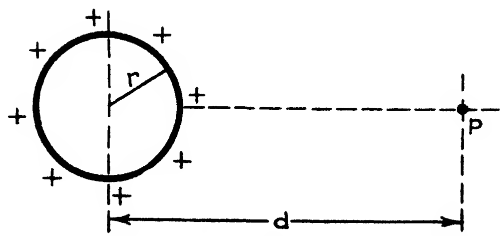


FIG. 52.—An insulated sphere exhibits a capacitance numerically equal to its radius.

at some point p outside the sphere will be Q/Kd, where d represents the distance of p from the center of the sphere. At the surface of the sphere $d = r$, and the potential there would accordingly be Q/Kr. This may be written

$$Kr = \frac{Q}{V}.$$

Now Q/V by definition [Eq. (71)], is numerically equal to capacitance; hence

$$Kr = C.$$

Therefore in free space

$$C = r. \tag{72}$$

Thus we see that the capacitance of an electrically isolated spherical conductor in free space is **numerically** equal to its radius. In the cgs system the radius would be expressed in centimeters; hence **the cgs electrostatic unit of capacitance is the centimeter.** This unit is frequently referred to as the **statfarad.** A sphere thoroughly insulated and well removed from other charged bodies may be used as a standard of capacitance.

57. Units of Capacitance. In the preceding section it was shown that the electrostatic unit of capacitance turned out to be a unit of length, and that in the cgs system this unit would naturally be the centimeter, or statfarad. Equation (71) would, then, be expressed thus,

$$C \text{ (statfarads)} = \frac{Q \text{ (statcoulombs)}}{V \text{ (statvolts)}}.$$

If, however, Q is expressed in coulombs (3×10^9 statcoulombs) and V

in volts ($\frac{1}{300}$ statvolts) the ratio Q/V would give

$$\frac{3 \times 10^9}{\frac{1}{300}} = 9 \times 10^{11}.$$

If, then, we had a unit of capacitance equal to 9×10^{11} esu (statfarads) it would be possible to express all three of the factors in Eq. (71) in terms of engineering units. The unit which answers this purpose is known as the **farad**; it is equal to 9×10^{11} statfarads (esu or centimeters). Equation (71) accordingly becomes

$$C \text{ (farads)} = \frac{Q \text{ (coulombs)}}{V \text{ (volts)}},$$

and we may say that a conductor has a capacitance of one farad if a charge of one coulomb gives it a potential of one volt. But it so happens that the farad is a magnitude of capacitance which is rather large for common use; hence a convenient fraction of this unit is ordinarily employed. The smaller unit has a value of 1 millionth of a farad; it is known as the microfarad. On this basis

$$1 \text{ microfrad} = 9 \times 10^5 \text{ statfarads.}$$

In certain lines of communication engineering, where capacitance values of very small magnitudes are encountered, a still smaller practical unit is sometimes employed. This is known as the micromicrofarad; it is equal to 1 millionth of a microfarad, or

$$1 \text{ micromicrofarad} = 10^{-12} \text{ farad} = 0.9 \text{ statfarads.}$$

Microfarad is usually abbreviated as μf, and micromicrofarad as $\mu\mu$f. To summarize

$$\text{Farads} = \frac{\text{esu}}{9 \times 10^{11}} = \frac{\text{cm}}{9 \times 10^{11}} = \frac{\text{statfarads}}{9 \times 10^{11}}$$

$$\text{Microfarads} = \frac{\text{esu}}{9 \times 10^{5}} = \frac{\text{cm}}{9 \times 10^{5}} = \frac{\text{statfarads}}{9 \times 10^{5}}$$

$$\text{Micromicrofarads} = \frac{\text{esu}}{0.9} = \frac{\text{cm}}{0.9} = \frac{\text{statfarads}}{0.9}$$

$$\text{Farads} = \frac{\text{microfarads}}{10^6} = \frac{\text{micromicrofarads}}{10^{12}}$$

$$\text{Microfarads} = \text{farads} \times 10^6.$$

It is important that the student become familiar with the several units of capacitance, and their interrelations.

Problem. A conductor having a capacitance of 500 $\mu\mu$f is found to have a potential of 3 statvolts. What is the magnitude of the charge on the body?

Solution. Equation (71) may be changed to the form $Q = CV$; all units involved must be reduced to a common type. This gives

$$Q = 500 \times 0.9 \times 3 = 1{,}350 \text{ statcoulombs}$$
$$= \frac{1{,}350}{3 \times 10^9} = 4.5 \times 10^{-7} \text{ coulomb.}$$

A second solution would be

$$Q = 500 \times 10^{-12} \times 3 \times 300 = 4.5 \times 10^{-7} \text{ coulomb.}$$

58. Capacitance of Two Concentric Spheres. In Sec. 57 we examined the case of a single electrically isolated body. Let us now consider the influence of a neighboring body. Suppose we have a second sphere **concentric** with the first, and grounded as shown in Fig. 53. Let the inner sphere carry a positive charge of Q units. Under these circumstances the inner surface of the outer sphere will manifest an equal negative charge. According to the deductions arrived at in Sec. 12, the field intensity at a point p due to this negative charge will be zero. However, the field strength at p due to the charge $+Q$ at any point will be [Eq. (9)] Q/Kx^2, where x is the distance to the point p. Therefore the potential difference between the two charged surfaces will be

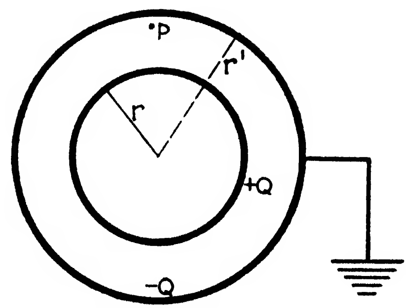


FIG. 53.—Two concentric spheres, one of which is grounded, act as a condenser.

$$V = -\int_{r'}^{r} \frac{Q}{Kx^2}\,dx = \frac{Q}{Kr} - \frac{Q}{Kr'} = \frac{Q(r' - r)}{Krr'} \qquad \text{statvolts.}$$

But since $C = Q/V$ we have

$$C = K\frac{rr'}{r' - r}, \tag{73}$$

as an expression giving the capacitance of two concentric spheres, the outer of which is at earth potential. If the radii are expressed in centimeters C will be in statcoulombs (centimeters). A comparison of Eqs. (72) and (73) shows that the presence of a second and oppositely charged conductor greatly enhances the capacitance of a simple spherical body. It should also be noted that the capacitance in this and other similar cases will depend upon which component is grounded. Equation (73) **will not give the capacitance** if the inner sphere is the grounded one. (Do you see why?)

Returning again to the two spheres above considered, as the radii of the concentric spheres are given larger and larger values, all small corresponding sections of the surfaces tend to become parallel planes. If r and r' are nearly equal, and if we let t represent $r' - r$, *i.e.*, the thickness of the dielectric, we may write

$$C = \frac{Kr^2}{t}.$$

If now we multiply both numerator and denominator by 4π, we get

$$C = \frac{4\pi r^2 K}{4\pi t}.$$

But $4\pi r^2$ is the area of the sphere; hence

$$C = \frac{KA}{4\pi t}. \tag{74}$$

where A is the effective area of one of the two surfaces. If A and t are in centimeters, C will be in centimeters, or statfarads. Thus, indirectly, we have deduced an expression giving the capacitance of two parallel plates.

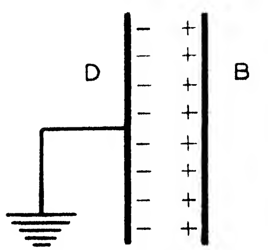


FIG. 54.—The potential of a charged plate is lowered by the presence of a second grounded plate, thus augmenting the capacitance of the first plate.

At this stage of our discussion concerning capacitance we may well ask ourselves the question: Why does the presence of a second earthed conductor tend to lower the potential of the body originally considered?

In Fig. 54, suppose that we have a plate B charged positively as shown. The potential of B would be equal to the amount of work required to bring unit positive charge from infinity to the plate, and this work, in turn, would be determined by the field strength due to the charge on B.

If now we place a second conductor D near to B and connect it to earth a negative charge will be induced on this second conductor and, because of this negative charge, **the resultant electric intensity at all points in the field will be less than formerly.** (Do you see why?) This means that the work required to bring our test charge from infinity, under the changed conditions, will be less, and hence it must follow that the potential of B has been lowered by the presence of the grounded conductor. It follows, therefore, that the capacitance will be increased [Eq. (71)].

59. Capacitance of Parallel Plates. The result expressed by Eq. (74) may be attained by another analytical procedure. Assume a posi-

tively charged plate B (Fig. 54) located near a second plate D, the latter being connected to earth. As indicated above, a negative charge will appear on D. The field intensity at any point between the plates (except very near the edges) will be $4\pi\sigma/K$, where σ is the charge per unit area on one of the plates. The work done in moving unit test charge from one plate to the other would measure the potential difference between the plates. Hence we may write

$$W = V_D - V_B = -\int_0^t \frac{4\pi\sigma}{K}\,dl = \frac{4\pi\sigma t}{K} = V_B$$

where t is the thickness of the dielectric. But $Q = \sigma A$, where A is the effective area of one of the plates. The last equality is true because D, being grounded, is at zero potential; hence by substitution we get

$$C = \frac{KA}{4\pi t},$$

which is identical with Eq. (74). By introducing the necessary transformation factors the above equation becomes

$$C = \frac{KA}{1.131 \times 10^7 t} \quad \mu\text{f}, \tag{75}$$

where t is in centimeters and A in square centimeters.

Owing to the fact that the field near the edge of the plates is not uniform, the formulas just developed will not give highly accurate values. These expressions can be made to apply rigorously, however, if a "guard ring" similar to that used in the case of the Kelvin absolute electrometer (Sec. 47) is provided for one of the plates.

Problem. Two parallel circular plates are separated by glass ($K = 6$) which is 2 mm thick. The plates are 14.5 cm in diameter. What is the capacitance of the assembly in statfarads? In microfarads? In micromicrofarads?

Solution. Using Eq. (74) we have

$$C = \frac{6\pi(7.25)^2}{4\pi \times 0.2} = 394.2 \text{ statfarads.}$$

Employing Eq. (75),

$$C = \frac{6\pi(7.25)^2}{1.131 \times 10^7 \times 0.2} = 0.000438\ \mu\text{f}.$$

Changing the result in statfarads to micromicrofarads,

$$\frac{394.2}{0.9} = 438\ \mu\mu\text{f}$$

60. Capacitance of Coaxial Cylinders. Referring to Fig. 55, assume two long coaxial cylinders, the radius of the inner one being r and that

of the outer r'. Let $+q$ be the charge per unit length on the inner cylinder. The equivalent induced negative charge on the inner surface of the outer cylinder must be taken into consideration, as in previous cases. The negative charge will not contribute (Sec. 13) to the field intensity at a point in the region between the two cylinders. The field intensity at any such point will accordingly be given by $2q/Kx$,

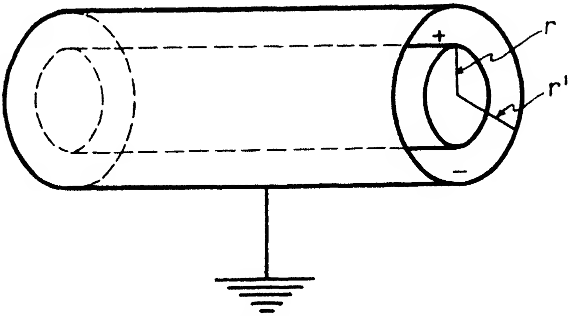


FIG. 55.—Coaxial cylinders, such as submarine and coaxial cables, act as condensers.

where x is the distance from the surface of the inner cylinder to the point being considered. The work done in moving unit test charge from the outer to the inner cylinder will be

$$W = -\int_{r'}^{r} \frac{2q}{Kx}\,dx,$$

and this, by definition, is equivalent to the potential difference between the two surfaces. Since the outer cylinder is grounded, the above expression will be the potential of the inner cylinder; hence

$$V = -\frac{2q}{K}\int_{r'}^{r}\frac{dx}{x} = \frac{2q}{K}\log\frac{r'}{r}.$$

Transposing and making use of Eq. (71)

$$C = \frac{K}{2\log_e r'/r} \qquad \text{statfarads/unit length.} \tag{76}$$

It is to be noted that here is a mathematical case in which the logarithm of a number **is used as a number.** Also, when using the above relation in calculations, it should be remembered that the logarithmic term in the denominator is the **natural logarithm** having the base $e(= 2.718)$. Since we may transform natural logarithms to logarithms having 10 as a base by multiplying by 2.3026, we have, for a cylinder of any length l,

$$C = \frac{Kl}{2(\log_{10} r'/r)2.3026 \times 9 \times 10^5} \qquad \mu f, \tag{77}$$

where l is in centimeters. Since r'/r is a ratio, the units in which they are expressed is immaterial, as long as they are both in the **same** units.

The above relation is important because it serves as a basis for the calculation of the capacitance of submarine cables. In such a case, the surface of the metallic conductor acts as the inner cylinder and the water as the outer cylinder. The insulating sheathing serves as the dielectric, the constant being, in cable practice, of the order of 3. Since

$$1 \text{ mile} = 160{,}934.4 \text{ cm},$$

and since the ratio

$$\frac{160{,}934.4}{2 \times 2.3026 \times 9 \times 10^5} = 0.0388,$$

we may write for the capacitance of a cable, in microfarads per mile,

$$C = \frac{0.0388K}{\log_{10}(r'/r)}. \tag{78}$$

Problem. A submarine cable consisting of a stranded conductor as the core and gutta-percha insulation is 2,200 miles long. The mean diameter of the core is 250 mils and the diameter outside the insulation is 500 mils.[1] If the gutta-percha has a dielectric constant of 3.6, what is the capacitance of the cable?

Solution. Substituting in Eq. (78), we have

$$C = \frac{0.0388 \times 3.6 \times 2{,}200}{\log_{10} {}^{250}\!/_{125}} = \frac{0.0388 \times 3.6 \times 2{,}200}{0.3010} = 1{,}020.6 \ \mu f.$$

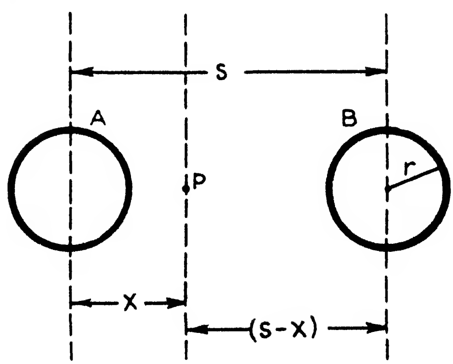


FIG. 56.—Parallel conductors, such as telephone, telegraph, and power wires, have capacitance and hence act as condensers.

61. Capacitance of a Pair of Telephone, Telegraph, or Power Wires. Let Fig. 56 represent a sectional view of the two wires of a telephone or power transmission line. In order to determine the capacitance of such

[1] The "mil," which is one-thousandth of an inch, is frequently employed as a unit of length in electrical engineering practice.

a pair of wires it will first be necessary to find the potential difference between the wires due to a positive charge on one and a negative charge on the other. Consider a point p on a line joining the centers of the two wires A and B. The field strength at p will be the sum of the intensities due to $+q$ on one wire and $-q$ on the other, where q is the charge per unit length. It has previously been shown (Sec. 13) that the intensity at any point due to the charge on a uniformly electrified cylinder is given by the expression $2q/x$, where x is the distance from the axis of the cylinder to the point in question. In this case the total field intensity at any point p will be

$$\mathcal{E} = 2q\left(\frac{1}{x} + \frac{1}{s - x}\right)$$

The work done in moving unit positive charge from center to center would give the difference in potential between the centers. This would be

$$W = V_A - V_B = -2q\int_{x=r}^{x=s-r}\left(\frac{s}{x(s - x)}\,dx\right)$$

Integrating we get

$$V_A - V_B = 4q \log_e \frac{s - r}{r} \qquad \text{statvolts}$$

In general, r is small compared with s; hence the above equation reduces to

$$V_A - V_B = 4q \log_e \frac{s}{r}$$

This leads to

$$\frac{q}{V_A - V_B} = C = \frac{l}{4 \log_e (s/r)} \qquad \text{statfarads/unit length.} \tag{79}$$

For any length l of a two-wire line we may then write, as the capacitance in microfarads,

$$C = \frac{l}{4 \times 2.3026 \times 9 \times 10^5 \log_{10} (s/r)} = \frac{0.0000001208l}{\log_{10} (s/r)}, \tag{80}$$

l being in centimeters, and s and r in the same units.

Since one mile = 160,934.4 cm, the capacitance, in microfarads per mile, of two parallel wires is

$$C = \frac{0.0194}{\log_{10} (s/r)}. \tag{81}$$

In deriving this relation we have assumed, for the sake of simplicity, that the wires were far enough apart so that the mutual interaction

of the charges would not disturb the uniform perimetrical distribution of the charge. We further assumed that the wires were high enough from the earth to preclude any influence from that source. In practice, neither of these ideal conditions obtains; hence the above formula yields only approximate numerical results. The measured values are, in general, higher than the computed values, the difference depending upon local conditions. However, the relation is useful in connection with communication and power engineering. The capacitance of communication and power lines plays a very important part in the electrical functioning of such conductors.

Problem. What is the capacitance of a pair of telephone wires 1 mile in length, the size of the wire being No. 10 B & S, and the spacing 1 ft?

Solution. One foot = 30.48 cm; No. 10 B & S wire has a diameter of 0.2588 cm. Substituting in Eq. (81),

$$C = \frac{0.0194}{\log_{10} 30.48/0.1294} = 0.008176\ \mu f.$$

62. Condensers. We have already seen (Sec. 16) that the medium surrounding a charged body is a seat of potential energy, and that the magnitude of the energy content is a function of the field intensity [Eq. (14)]. It was also shown that the value of the field intensity depends upon the magnitude of the charge involved. It therefore follows that the energy content of the medium will be a function of the magnitude of the charge giving rise to the field. Therefore, if the charge on a conductor can be increased the energy thus stored in the medium will be augmented. But we have also seen (Sec. 41) that as the charge on a body increases its potential rises. This increase in potential, if carried very far, may result in a breakdown of the dielectric, with a consequent dissipation of the stored energy. If, therefore, we are to store electrical energy by charging a conductor, it becomes important, if possible, to keep the potential from rising seriously as the charge is augmented. Changing Eq. (71) to the form $V = Q/C$ it is evident that, for a given charge, any circumstance that would **increase** the capacitance would **decrease** the corresponding potential. It therefore follows that the energy-storing ability of a conductor will be increased if means can be found to increase its capacitance. We have already seen (Sec. 59) that the capacitance of a conducting plate is greatly increased by associating with it a second grounded plate. This, or any other, arrangement of conductors whereby the capacitance of a single conductor may be augmented is known as **a condenser.** In general, condensers consist of one or more pairs of electrical conductors separated by some form of dielectric. The pre-

viously discussed concentric spheres and coaxial cylinders in reality constitute condensers, and are, in fact, both used as standard capacitance units because their mechanical dimensions can be accurately measured. But the most common form of condenser consists of one or more pairs of parallel plates, as discussed in Sec. 59. Indeed, Eqs. (74) and (75) are applicable to this type of unit.

The condenser is one of the oldest known pieces of electrical apparatus and is, in its modern forms, in extensive use. It is said that the principle of the condenser was originally discovered by a German named von Kleist in 1745. The device used by Kleist consisted of a glass receptacle filled with water and held in the hand. Kleist's experiment was repeated by Cuneaus at Leyden and the early form of condenser came to be known as a "Leyden jar." A few years afterward Benjamin Franklin designed condensers having tin-foil coatings pasted on both glass jars and glass plates. The type of condenser developed by Franklin continued to be used with little if any change in form for something like 160 years. It was not until the advent of radiotelegraphy that improved types of condensers came to be produced. The original Leyden jar, introduced by Franklin, has in recent years been replaced by compact, low-loss units having capacitance values ranging from a few micromicrofarads to several hundred microfarads. Solids, liquids, and gases are used as dielectric mediums. Any medium employed as a dielectric must possess several important characteristics. These properties are discussed in Sec. 64.

Most condensers which are to be operated at potentials ranging from 500 to about 25,000 volts are made up of alternate layers of very thin metal and carefully tested mica, the whole unit being compressed and impregnated *in vacuo* with a suitable nonhygroscopic insulating compound. Mica is used because it has a high dielectric constant, a relatively high dielectric strength, and can easily be split into very thin sheets. Condensers which are to be used as secondary standards of capacitance usually are of the mica-dielectric type. Condensers built to operate at high potentials and high frequencies have capacitance values of the order of a few hundred micromicrofarads. Recently, manufacturers have reverted to an old practice in the construction of certain condensers used in radio communication. The conducting components of these condensers consist of a metallic coating of silver or other metal plated on the surface of the dielectric, which is either mica or some ceramic. Condensers so constructed have a low and definitely known temperature coefficient, that in some cases is actually zero. Such units are electrically stable and very compact.

Condensers consisting of alternate layers of thin metal foil and impregnated paper, commonly made by rolling together long strips of the foil and paper, are widely used in telephone and telegraph practice. Such con-

densers are commonly made in units of 0.5 to 2 μf and are designed to be operated at potentials of a few hundred volts.

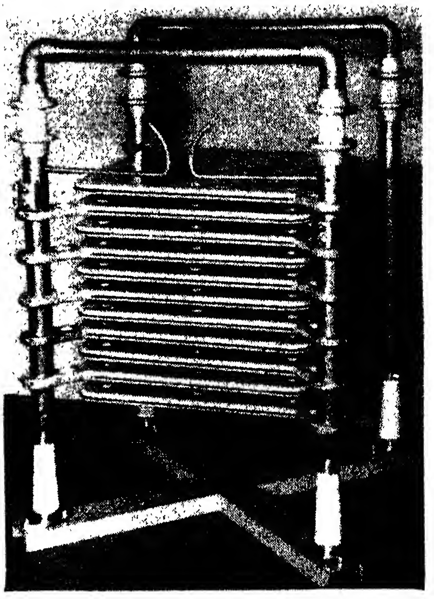


FIG. 57a.—A fixed air condenser—high-potential type. (*E. F. Johnson Co.*)

In radio and guided-wave (carrier current) practice air condensers are extensively employed. One advantage of this type of condenser is that the capacitance of a given unit may be made variable. Another advantage is that the dielectric is "self-healing" in the event of rupture. When a condenser is charged by an alternating potential the dielectric is subjected to an alternating physical stress, and this results in an appreciable dissipation of energy in the form of heat. Gases have low dielectric hysteresis loss (Sec. 64) and it is chiefly for this reason that both fixed and variable air condensers are widely used in radio transmitting and receiving equipment. Air condensers are made in sizes ranging from a few micro-microfarads to something of the order of 500 $\mu\mu$f. Figure 57 shows a fixed and a variable air condenser designed for use in radio transmitters.

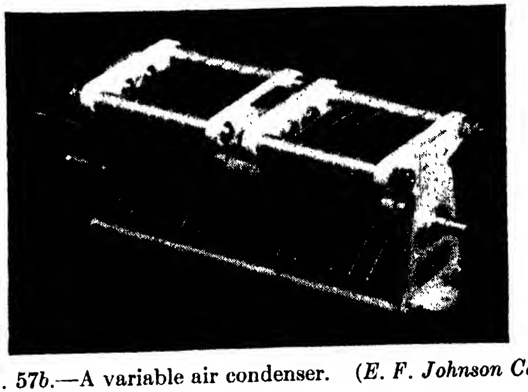


FIG. 57b.—A variable air condenser. (*E. F. Johnson Co.*)

Variable condensers are of several types, depending upon the shape of the movable plates. If it is desired to have a condenser whose capacitance varies directly as the angular displacement of the movable plates they are made semicircular in form.

When a variable condenser forms a part of an electrical wavemeter, the use of semicircular plates results in a wave-length scale that is non-uniform, being crowded too closely at the lower readings. To avoid this, it becomes necessary, for reasons which we shall see later, so to design the movable plates that the capacitance shall vary as the **square** of the displacement.

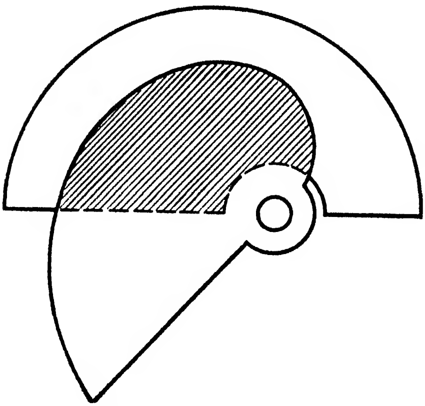


FIG. 58.—Shape of condenser plates to give a change in capacitance proportional to the angular displacement.

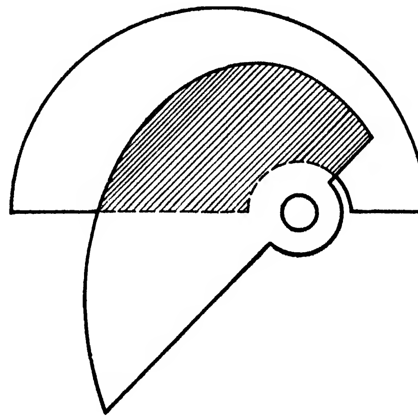


FIG. 59.—Shape of condenser plates which will give the same percentage change in capacitance for each scale division.

In order that the effective area between the fixed and movable plates shall vary as the square of the angle of rotation, the movable plates must have a shape, the boundary curve of which is given by the equation

$$r = \sqrt{4a\phi + r_2^2},$$

where r is the radius of the movable plate; r_2 the radius of the circular area which must be cut from the fixed plate in order to provide clearance for the shaft supporting the movable plates; and a is a constant relating ϕ to the capacitance. Figure 58 shows the approximate contour of the plates of such a condenser.

In certain h-f electrical measurements it becomes necessary to have available a type of variable condenser that, for a given angular displacement, will produce the same **percentage** change in capacitance at all parts of the scale. This condition will obtain if the boundary curve of the movable plates is given by the equation

$$r_1 = \sqrt{2C_0\alpha\epsilon^{\alpha\phi} + r_2^2},$$

where C_0 represents the capacitance when the angular displacement ϕ is

zero; α (a constant) the percentage change in capacitance per scale division; and r_2 has the same meaning as in the previous case. A pair of plates conforming to these conditions is shown in Fig. 59. For an authoritative discussion of air-condenser design the student is referred to the *Bureau of Standards, Circular* 74, the section on air condensers.

If a fixed condenser of the gas-dielectric type is to be subjected to high potentials, the plate assembly is sometimes inclosed in a gastight housing and operated at a gas pressure of several atmospheres. Both the dielectric strength and the dielectric constant have greater values at the higher pressures (see table, page 101).

The utilization of a gas as a dielectric in high-potential condensers necessitates relatively wide separation of the plates and careful rounding of their edges, the latter to avoid a brush discharge. The insulating bushings must be of the highest grade of insulating material.

Oil is occasionally used as a dielectric in condenser operation. Its dielectric strength and its constant are greater than for air, and in addition it is self-healing. However, if breakdown occurs, the oil will be partially carbonized and thus become conducting. Transformer and mineral oils are suitable for use in condensers.

The **electrolytic condenser** is a special type of capacitance unit which is extensively used at the present time. The so-called "dry" electrolytic unit is the type which is most widely employed. These units are of the "rolled" type and consist of two strips of aluminum foil separated by a layer of gauze that is saturated with the electrolyte. Each manufacturer has his own particular electrolyte, but one commonly used consists of a mixture of boric acid, glycerine, and ammonia water. The components are tightly compressed and then subjected to a d-c potential of several hundred volts. As the result of an electrochemical reaction, a very thin oxide film is formed on the aluminum strip that is connected to the positive side of the applied potential. The film thus formed is a nonconductor and therefore serves as a dielectric layer. The electrolyte functions as the other element of the condenser, the second aluminum strip simply serving to make electrical contact with the electrolyte. Owing to the extreme thinness of the insulating layer [t in Eq. (74)] units of large capacitance values can be had in very small sizes. They are made in units ranging from a few to several thousand microfarads. If potentials in excess of about 500 volts are applied to this type of unit, the oxide-insulating anode film will break down, and it will temporarily cease to function as a condenser. However, it is possible to cause the oxide film to reform and thus repair the damage. Electrolytic condensers show a greater electric leakage than do those employing solids or gases as dielectrics.

63. Energy Content of a Charged Condenser. Reference has already been made to the fact that condensers are utilized for the purpose of storing or absorbing electrical energy. It will therefore be convenient to have available an expression which will give the energy content of a charged condenser in terms of its capacitance and the potential to which it is charged.

It has already been shown [Sec. 46, Eq. (65)] that the work done in charging a body is given by the relation

$$W = \tfrac{1}{2}QV.$$

This equation is equally applicable to a charged condenser, in which case Q will be the charge resident in the condenser, and V the potential difference between the two conducting elements. But Eq. (71) shows that

$$Q = CV;$$

hence

$$W = \text{P.E.} = \tfrac{1}{2}CV^2. \tag{82}$$

If C is in statfarads and V in statvolts, the energy will be in ergs. Remembering that a farad is equivalent to 9×10^{11} statfarads, and that a volt represents $\frac{1}{300}$ statvolt, one may express the terms in the above equation in engineering units. Introducing these factors on the right-hand side of our equation, we get

$$\frac{CV^2 300^2}{9 \times 10^{11}} = \frac{CV^2}{10^7}.$$

To preserve equality we must divide W by 10^7; this reduces our result to joules. Hence we may write

$$W \text{ (joules)} = \tfrac{1}{2}C \text{ (farads) } V^2 \text{ (volts)}. \tag{83}$$

By a similar procedure it may be shown that

$$W \text{ (joules)} = \tfrac{1}{2}\,\frac{C(\mu\text{f})\,V^2 \text{ (volts)}}{10^6}. \tag{84}$$

Problem. What is the energy content of a condenser having a capacitance of 500 $\mu\mu$f when charged to 10,000 volts?

Solution. Substituting an Eq. (83), we have

$$W = \frac{500 \times (10{,}000)^2}{2 \times 10^{12}} = 0.025 \text{ joule}.$$

64. Dielectrics. From what has been said thus far it is evident that the medium separating the plates of a condenser must possess certain electrical characteristics if it is to function effectively. Among these characteristics may be mentioned

1. High dielectric constant
2. High dielectric strength
3. Low dielectric loss
4. High insulation resistance
5. Low temperature coefficient
6. High tensile and compressive strength

Space and time will not permit a detailed discussion of the above-mentioned characteristics, but we may at least note the general significance of the several points listed.

DIELECTRIC CONSTANTS

	K	Dielectric strength[1]
Air (normal pressure, 0°C)	1.000590	30
Air (20 atmospheres, 19°)	1.0108	500
Carbon dioxide (normal pressure, 0°)	1.000985	36
Carbon dioxide (20 atmospheres, 15°)	1.020	?
Hydrogen (normal pressure, 0°)	1.000264	26.1
Alcohol, ethyl, 20°	25.8	?
Oil, transformer	2.5 ±	75 ±
Amber	2.9	?
Casein, moulded	6.4	?
Cordierite ceramics	5–5.5	100*
Empire cloth	?	80–300
Glass	5.4–8.5	300–1500
Magnesium titanate ceramics	12–18.	150*
Mica	3.0–5.95	300–2200
Paper, telephone, treated	2.5–4.	200–250
Paraffin, solid	2.1–2.46	250–450
Porcelain	5–7.5	200–280*
Quartz, fused	4.2	150–200
Rubber, hard	2–3	160–500
Resin	2.48–2.57	?
Steatites (magnesium silicate, etc.)	5.5–7.5	200–300*
Shellac	2.5–4.	?
Titanium dioxide ceramics	70–90.	100*
Titanium-zirconium dioxide ceramics	40–60.	150*

[1] Values not marked are given in kilovolts per centimeter; those marked *, in volts per mil. Breakdown values depend to some extent on the thickness of the sample; hence one cannot transform one value into an equivalent on another thickness basis.

In the case of condensers the necessity of a high K value is obvious; the higher the value of K the smaller can be the physical size of the con-

denser as a mechanical unit. This is an extremely important factor in many modern applications, particularly in connection with portable radio equipment and as components in hearing-aid assemblies. An examination of the table on page 101 will disclose the fact that certain recently produced ceramic dielectrics show remarkably high K values. This is particularly true of those mediums into the composition of which titanium dioxide enters. In one or two such cases K has a value of the order of 100.

It has already been pointed out (Sec. 8) that a dielectric which is subjected to an electrostatic field undergoes an actual internal (molecular) distortion, and if the intensity is sufficiently high, physical rupture of the medium will result. As the field strength is increased electrons are probably actually detached from their normal molecular association and caused to move **through the medium** toward the positive electrode. Thus, instead of having what we originally called a displacement current, there now exists **an actual transfer of electrons between the two electrodes.** In short, we have an electric current in the ordinary sense of the term. This transfer of electrons is apt to occur suddenly—a disruptive discharge takes place that in turn results in a physical rupture of the dielectric.

The ability of a dielectric to withstand physical rupture when subjected to electric stress is known as **dielectric strength** and is expressed in terms of field intensity. In Sec. 43 we saw that field intensity might be expressed [Eq. (59)] in terms of potential gradient. That relation is particularly useful here. It is customary to give the dielectric strength in volts (or kilovolts) per centimeter, millimeter, or mil, as indicated in the previous table.

In the practical use of insulating materials it is obviously important to have a knowledge of the field strength they can stand before rupture occurs, and also of the factors which tend to determine those values. The dielectric strength of a given specimen depends upon several factors, the most important of which are temperature, thickness of sample, time rate of applying the field, whether a continuous or alternating field is applied, and, if alternating, the frequency. The shape and size of the electrodes also have a bearing on dielectric strength.

In general, increase in temperature tends to lessen the dielectric strength. There is reason for believing that the electrons which figure in whatever conductivity obtains in the case of insulators are liberated by heat; hence such an effect might be expected. If the temperature is raised to such a point that the physical characteristics of the medium are modified, the dielectric strength will of course be decidedly changed.

In many cases the dielectric strength of a medium does not vary directly as the thickness of the test specimen. Commonly the dielectric

strength shows a smaller value for relatively thick samples than for those which are thin. The nonlinear relation which obtains between dielectric strength and thickness is probably due to the fact that the thicker the specimen the less homogeneous it is. However, thinly laminated materials give better values than nonlaminated samples of the same total thickness.

The length of time during which the field is applied to the sample has a marked effect on the results. In general a specimen will withstand decidedly higher potentials for periods of the order of a second or less than for materially longer intervals. This phenomenon is probably associated with the electronic displacements referred to earlier in this discussion.

If the applied voltage is alternating, the dielectric strength of a given material will usually be less than when tested with direct potential. The values in the latter case may run as high as two times the former. The frequency of the applied potential (in the region of audio frequencies) does not appear to figure appreciably, but at the extremely high frequencies used in communication engineering practice the dielectric strength is less than at ordinary commercial frequencies.

In case of solid insulating materials, particularly when in the form of thin sheets, it has been found that the dielectric strength is higher when the test is made with small electrodes than when larger terminals are employed. However, if the electrodes are needle points the reverse is true.

Reference was made above to the electronic displacements that occur when a dielectric is subjected to an alternating electrostatic field. These alternating electronic displacements constitute, in reality, an alternating electric current. As we shall see later, the existence of a current involves losses in the form of heat. Under certain circumstances the heat generated by the displacement current existing in the dielectric of a condenser may result in a marked rise of temperature, thus changing both the mechanical and electrical properties of the dielectric.

In addition to the thermal effects above mentioned, there are other dielectric losses which manifest themselves as heat. The alternating polarization of the atoms of the dielectric when subjected to an alternating field probably results in a physical distortion of the molecular structure which in turn gives rise to what amounts to an intermolecular frictional effect. That some such electromechanical change takes place in a dielectric is evidenced by the fact that when a potential difference is impressed upon the plates of a condenser the first rush of charging current is followed by a small but decreasing flow. A measurable length of time is thus required in order to establish a stable electrical state. A similar

phenomenon is manifest when a condenser is **discharged.** The charge and discharge of the condenser lag somewhat behind the applied potential. An energy loss is associated with this phenomenon. Just what happens within the dielectric is not definitely known, but we do know that a **condenser does not give up all of the energy which it acquires during the charging process.** This loss of energy is referred to as **dielectric loss.** Its magnitude, in any given case, depends upon the nature of the dielectric involved, but does not depend upon the applied potential of the frequency. Air shows practically no dielectric loss, and the new ceramic dielectrics have very little loss of this character. In general, a rise in temperature results in an increase in the dielectric-loss factor. The expression **dielectric hysteresis** is sometimes used as a term to cover the frictional and lag effects above mentioned. Strictly speaking, it should apply only to the lag phenomenon. The losses associated with the intermolecular friction and the lag effect cause a condenser to have a power factor (Sec. 168) that is not zero. The magnitude of this power factor is, in fact, commonly taken as an index of the dielectric losses in any given case. The power factor for air is zero.

Like some other phenomena as, for example, ordinary mechanical friction, dielectric losses have recently been put to extensive practical use. The heating of dielectrics (nonconductors) by the application of h-f alternating fields is currently being employed in connection with the fabrication of laminated wooden structures, the dehydrating of foods, and in other related fields. In producing such thermal effects both the I^2R heat due to the displacement currents and the heat due to the dielectric hysteresis is made use of. It is probable that this form of localized heating will be greatly extended in the near future.

In order to function effectively as a dielectric for use in condensers and insulators a medium must possess a high and permanent resistance value. Not only must the body resistance of the dielectric be high (of the order of 100 megohms) but its **surface resistance** must also be high. The author has found that some widely used synthetic dielectric materials often show very low insulating values, due apparently to surface leakage —even in air at ordinary humidity conditions.

Surface leakage may become a serious matter in connection with the use of insulating bushings; and the losses thus entailed, when added to possible dielectric losses, may become prohibitive.

Reference has already been made to the fact that with the new ceramic dielectrics, the temperature coefficient of condensers may be made either positive or negative and even zero. As we shall see later, condensers are frequently associated with other components in the design of electronic equipment. By the use of condensers having a negative temperature

coefficient the change in capacitance due to temperature changes may be made to compensate for positive temperature effects in other components. The capacitance changes per degree centigrade for these new mediums range from $+5 \times 10^{-4}$ to -6.5×10^{-4}.

65. Capacitance of Condensers in Series and Parallel. It frequently becomes necessary to design condensers for use at potential differences which are much higher than the dielectric strength of the available insulating material. In order to meet this situation, it is customary to connect several condenser units in series, thereby distributing the potential difference between the several units which go to make up the condenser as a whole. The danger of a breakdown of the dielectric is thus minimized, but by such an arrangement of condenser components the total capacitance is less than the capacitance of the individual units which

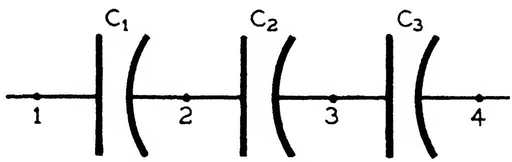


FIG. 60.—Condensers connected in series.

compose the condenser assembly. It is important to be able to compute the total capacitance of a group of condenser units connected in this manner.

In Fig. 60, let C_1, C_2, and C_3 represent the values of several condensers connected in **series.** Our problem is to find an expression for the total capacitance in terms of the capacitance of the individual units. By the laws of electrostatic induction, if these condensers are charged by connecting the points 1 and 4 to a source of potential difference, each condenser in the series will acquire an equal charge that we may call Q. Designating the potentials at the points 1, 2, 3, etc., by V_1, V_2, V_3, etc., we may write [Eq. (71)]

$$V_1 - V_2 = \frac{Q}{C_1},$$

$$V_2 - V_3 = \frac{Q}{C_2},$$

$$V_3 - V_4 = \frac{Q}{C_3}, \text{ etc.}$$

For the total difference of potential between 1 and 4 we would have

$$V_1 - V_4 = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots\right).$$

But $Q/(V_1 - V_4) = C$, the total capacitance; therefore

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \text{ etc.} \tag{85}$$

It is thus evident that, as implied above, **the combined capacitance of condensers in series is less than the capacitance of any one of the units composing the series.** Because of this fact, condensers are sometimes connected in series not only in order to avoid breakdown, but also to secure a capacitance value **less** than that of any unit which may be readily available for some particular purpose.

In order to secure **larger** capacitance values than would be given by a single condenser, or by a group arranged in series as just described,

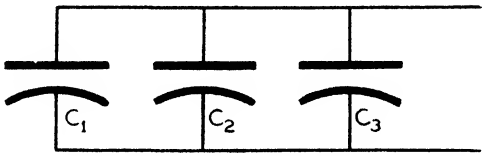


Fig. 61.—Condensers connected in parallel.

capacitance units may be connected in **parallel.** Let us now develop an expression for the total capacitance of a group of units connected in this manner.

A parallel arrangement is shown in Fig. 61. In this case all condenser units are charged to the same potential; hence, from our fundamental definition of capacitance [Eq. (71)], it is evident that each condenser unit will acquire a charge proportional to its individual capacitance, and the total charge will of course be equal to the sum of the charge on the several units. Hence we may write

$$Q = Q_1 + Q_2 + Q_3 + \cdots \text{ etc.,} \tag{i}$$

where Q is the total charge on the bank of condensers as a whole.

From Eq. (71) it follows that

$$Q_1 = C_1V, \qquad Q_2 = C_2V, \text{ etc.} \tag{ii}$$

Substituting the values given in (ii) in (i) and reducing, we have

$$C = C_1 + C_2 + C_3 + \cdots \text{ etc.,} \tag{86}$$

where C is the capacitance of the group of condensers as a whole. Thus we see that the total capacitance equals the sum of the capacitances of the several units.

In order to secure the desired capacitance, and at the same time keep down the potential difference across the elemental units, it is sometimes

necessary to assemble a series-parallel organization as shown in Fig. 62. In order to compute the capacitance in such a case one first proceeds to find the capacitance of each series group by the use of Eq. (85), and then deals with the values thus secured as a parallel combination by the aid of Eq. (86).

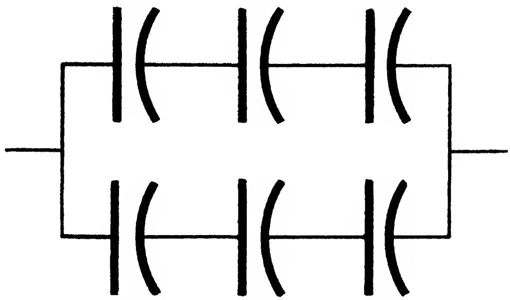


Fig. 62.—Condensers in a series-parallel combination.

Problem. Eight condensers are connected in a series-parallel arrangement similar to that shown in Fig. 62, four 1-μf units being in each series group. What is the total capacitance?

Solution. For each series group we would have

$$\frac{1}{C'} = 1/1 + 1/1 + 1/1 + 1/1 = 4.$$

The parallel arrangement of the two groups would then give

$$C = 1/4 + 1/4 = 1/2 \ \mu f.$$

66. Determination of Dielectric Constant. We have seen that the dielectric constant is an important factor in many calculations connected with electrostatics. It therefore becomes necessary to determine the value of this constant experimentally. Fortunately, this involves only a comparatively simple procedure, especially in the case of liquid and gaseous dielectrics.

An accurately made parallel-plate air condenser is assembled and its capacitance computed from its dimensions. The medium under test is then introduced between the plates, conditions being so arranged that it completely fills the space between the condenser plates. The capacitance is then experimentally determined by any convenient method, and the two values of capacitance compared. The theory involved is as follows.

In the case of air as the dielectric, we may write, using Eq. (74),

$$C_a = \frac{K_a A}{4\pi l}, \tag{i}$$

where C_a is the capacitance of the condenser when using air as a dielectric and K_a the constant for air.

Holding all mechanical factors constant, the situation with the material under test may be represented by the relation

$$C_x = \frac{K_x A}{4\pi t}, \tag{ii}$$

where C_x represents the capacitance in the second case and K_x the corresponding constant.

If now we divide (ii) by (i) there results

$$\frac{C_x}{C_a} = \frac{K_x}{K_a}.$$

For most purposes K_a may be taken as unity; hence

$$K_x = \frac{C_x}{C_a}. \tag{87}$$

The above expression shows that the dielectric constant of a specific medium is given by the ratio of the condenser capacitance with that medium as the dielectric to its capacitance with air as the medium. Indeed this ratio frequently serves as a definition of "specific inductive capacity," or dielectric constant.

PROBLEMS

1. A sphere 5 cm in diameter is charged to a potential of 10,000 volts. What is the magnitude of the charge involved? Express the result in statcoulombs and in coulombs.

2. The mean diameter of the earth is 7,918 miles. What is its capacitance in farads?

3. What work would be done in changing the potential of the earth by 1,000 volts? Express the result in joules.

4. Prove that the capacitance of a spherical condenser, with the inner conductor grounded, is given by the expression

$$\frac{R_1 R_2}{R_2 - R_1} + R_3,$$

where R_1 is the radius of the inner sphere, R_2 the inside radius of the outer shell, and R_3 the outside radius of the outer sphere.

5. A two-plate condenser of 0.00035 μf capacitance is to be constructed. Mica 0.06 mm thick is to be used as the dielectric, the K value being 5. What must be the area of each plate in square centimeters?

6. What conductor area would be required for a condenser of the same capacitance as that specified in the previous problem if a titanium dioxide ceramic 2 mm thick were to be used? Take an average value for K.

7. It is desired to construct a standard tubular condenser having a capacitance of 0.0002 μf. Tubing is available whose outside diameters are 5 and 4 cm, respectively. The wall is 2.5 mm thick. How long must the pieces of tubing be if we neglect the "fringing"?

8. It is desired to construct a mica condenser having an approximate capacitance of 1 μf. The mica has a thickness of 0.05 mm and has a constant of 5. The metal foil is so cut that the sheets will overlap 3×4 cm. How many sheets of metal and mica will be required?

9. An outside pair of telephone wires are size No. 19 (1 mm diameter). The wires are parallel, 100 ft in length, and separated 3 mm by insulating material whose dielectric constant is 2.5. What is the capacitance of the wires with respect to each other?

10. Two No. 4 wires are 50 miles long, spaced 3 ft apart. What will be the capacitance of the line? The diameter of No. 4 wire is approximately 0.6 cm.

11. Prove that the capacitance of a single wire of length l and radius r supported at a height h above, and parallel to, the earth's surface, will be given by the expression

$$C = \frac{1}{2 \log_e 2h/r} \quad \text{statfarads.}$$

12. A condenser of 300 $\mu\mu f$ is at hand, but it is desired to have available a condenser combination whose capacitance shall be 200 $\mu\mu f$. What must be the capacitance of the condenser which can be connected in series with the original unit in order to give the required capacitance?

13. It is desired to have a condenser unit whose capacitance shall be 0.4 μf; and that the unit shall be charged to a potential of 5,000 volts. Ten 2-μf condensers are available, each one of which will stand 1,000 volts potential difference across its terminals. How should the unit be assembled?

14. The potential difference between two wires of a high potential system is thought to be of the order of 10,000 volts. A supply of 1-μf condensers is available, each capable of withstanding 1,000 volts. An electrostatic voltmeter is also at hand whose full scale reading is 1,000 volts. How would you arrange and use this equipment in order to determine the actual potential difference between the line wires?

15. A charged parallel-plate air condenser shows a potential difference of 500 volts between its terminals. If a piece of insulating material having a K value of 5 was introduced between the plates and had such a thickness that it completely filled the space, by what amount would the potential be changed?

16. In the previous problem, was the force between the plates changed when the solid dielectric was introduced, and if so, how much? Was the energy content of the condenser changed, and if so, by how much?

17. A No. 12 wire 30 m in length is supported horizontally at a height 5 m above the earth's surface. What is its capacitance? No. 12 wire has a diameter of 0.2 cm.

CHAPTER VIII

CURRENT, ELECTROMOTIVE FORCE, AND RESISTANCE

67. The Electric Current. Thus far in our study we have been considering the case of charges **at rest.** We are now to examine those phenomena that manifest themselves when charges are in motion. Thus we enter the field of electrodynamics, in contradistinction to electrostatics. One or more electrons moving in a definite direction constitute what we shall speak of as an electric current. This orderly electronic motion may take place through a solid, a liquid, a gas, or in free space. In the case of liquids and gases, as we shall see later, the current is dual in character, consisting of charged portions of atoms. For the time being we will confine our attention to the case of currents in solids.

The **time rate** at which electrons pass any given cross section of a conductor determines the magnitude of what is referred to as **current strength.** Mathematically this concept may be represented thus,

$$i = \frac{dq}{dt}, \tag{88}$$

where i is the **instantaneous** value of the current strength. If the time rate of electronic motion (flow) is constant, we may say

$$I = \frac{Q}{t}, \tag{89}$$

where I is the **average** current strength and Q the total number of electrons passing some reference plane during the time t. From the last expression it follows that the total quantity of electricity which passes the reference plane will be given by the relation

$$Q = It.$$

The question of units at once presents itself. In what units are we to express current strength? In Eq. (89), if Q is in statcoulombs and t in seconds, I will naturally be in electrostatic units. Since the statcoulomb represents a very small quantity of electricity, the esu unit of current turns out to be too small for practical use. Another unit has therefore been adopted which is known as the **ampere** which, for the present, shall be taken as the equivalent of 3×10^9 esu. It follows then that if Q is in

coulombs and t in seconds, I will be in amperes. Later we shall encounter two other methods of defining the ampere.

In passing, it should be pointed out that two subdivisions of the ampere are in common use. In research and in engineering work, particularly in communication engineering, very small current magnitudes are often encountered. One of these units is the **milliampere,** which is 1/1,000 amp; the other is the **microampere,** equivalent to 10^{-6} amp.

68. Electromotive Force. In our study of electrostatics we saw that expenditure of energy is involved in the act of separating unlike charges. It was also noted that work is done (energy consumed) in moving a quantity of electricity between two points which are not at the same potential. The latter fact is embodied in Eq. (53), which should be

FIG. 63.—Electrons tend to move between two points that are not at the same potential.

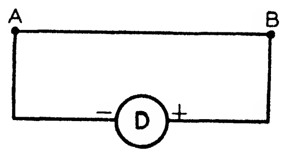


FIG. 64.—A source of energy is necessary in order to maintain a current.

re-examined at this point. In Fig. 63 let A and B represent two bodies not at the same potential. Suppose B to be at the **higher** potential. Under these circumstances, if A and B are electrically joined by means of a conductor electrons will move from A toward B, *i.e.*, **toward the point of higher potential.** In short, an electric current results from the fact that **a difference of potential exists between A and B.** If the above-indicated system is left to itself the transfer of electrons from A to B will, in time, cause B to attain the same potential as A, with the result that the electron flow will cease. In other words, our current has been a **transient** one.

If, however, we were to provide a more or less permanent source of electrons and a complete circuit around which they might move the situation would be different. Suppose we arrange conditions as shown in Fig. 64, where D represents, say, a dry cell or some other agent which is capable of maintaining a difference of potential between A and B. Under these circumstances the electrons will continue to flow from A to B and will, in fact, move completely around the circuit, passing through the "generator" D itself. Electrons will be supplied continually to A and abstracted at the same rate from B—this as a result of the functioning of the so-called "generator" D. Energy is required in order to move electrons from A to B and from B through D to A again. By Eq. (53) the

total work done in accomplishing this would be equal to $V_{AB}Q + V_SQ$, or $Q(V_{AB} + V_S)$; where V_{AB} represents the difference in potential which is being maintained between A and B, and V_S **the potential difference required to cause movement of the electrons through the generator itself and through the connecting wires.** Stated in another way

$$V_{AB} + V_S = \frac{\text{work done}}{Q}.$$

If, in the above relation, we make Q equal unity it could then be said that the quantity represented by $V_{AB} + V_S$ is equal to the work done **in moving the unit charge once around the entire circuit.** This work is done by D, and the energy thus involved is called the **electromotive force** (emf) of the source or generator. The last equation may now be written thus,

$$\text{Work done} = (\text{emf})\ Q = (\text{emf})\ It,$$

which, if we make I unity, may take the form

$$\frac{\text{Work done}}{\text{Time}} = \text{emf}. \tag{90}$$

In other words, **emf is the time rate at which a source supplies energy when unit current exists in the circuit.**

It will be recalled that potential and potential difference are expressed in terms of work done per unit charge, *i.e.*, in statvolts or volts. In view of the above definitions of emf it is natural, then, to express this quantity in the same units. If 1 joule of work is done in driving one coulomb of electricity once around the complete circuit, the emf of the source has a value of 1 volt.

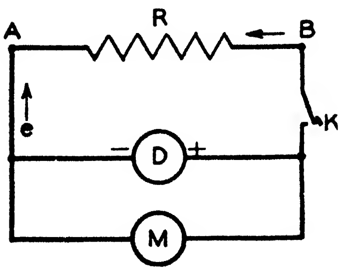


FIG. 65.—Illustrating why the terminal potential difference of a generator may be less than its emf.

In any discussion of emf there are one or two distinctions that should be carefully noted. In the first place, the potential difference **at the terminals of the generator** when it is supplying energy to the circuit is **not** the emf developed by that source. In Fig. 65, we see essentially the same circuit as that indicated by Fig. 64. There has been added, however, a connecting switch K and a meter M, of the electrostatic type, thus making it possible to determine the potential difference between the terminals of the generator D when it is **not connected** to the "load" circuit AB; and also when it **is** supplying energy to the load. By way of illustration, suppose that R represents the filament of a flashlight bulb and D a dry cell. With K open (no current flowing) the meter M would probably

read something like 1.5 volts. If now K is closed, thus connecting the lamp to the generator, it will be found that the meter will read, perhaps, 1.2 volts. Why the difference in reading? The **open-circuit reading gives the emf of the generator** (in this case the dry cell). The difference between the two readings (0.3 volt) is the potential difference V_S necessary to drive the electrons through the generator itself and through the connecting wires; and the 1.2 volts represents the work done V_{AB} in driving the electrons through the filament of the lamp. The potential difference at the terminals of any generator (battery, dynamo, or other source of electrical energy) **is always less on load than on open circuit**—and less by an amount equal to the fall in potential within the generator itself. This is a very important observation. The potential difference at the terminals of a generator when it is supplying energy to an external circuit is referred to as the **terminal potential.**

The term "fall in potential" was used above. In the external circuit (Fig. 65) there is a rise in potential as one travels with the moving electrons from A to B and **a fall** in potential (potential becomes less) if one were to move against the electronic stream, from B toward A, *i.e.*, in the direction of the conventional current. Likewise, in passing through the generator itself, there is a rise in potential as one passes from the negative to the positive terminal, and a corresponding fall in potential as one proceeds in the reverse direction. These two "drops," as they are called, are oppositely directed. This is in conformity with the observations made above.

Another point in the above connection is also worthy of note. We refer to the fact that a difference in potential (a "drop") exists between any two points in the external (load) circuit **only when current is flowing;** but the generator develops an emf, by chemical, mechanical, or other means, **even when not connected to a load circuit.**

Since work is done in moving the electrons through the conductor and through the generator the energy thus expended must appear in some form; and such is indeed the case. The expended energy appears as heat; both the conductor and the generator are heated. Later we shall see that there is a very definite relationship between the magnitude of the current and the quantity of heat developed. The amount of energy appearing as heat is independent of the direction of electron flow.

In the foregoing discussion we have dealt with the **electronic current,** and have noted that the flow is from points of low potential to points of higher potential. In other words, the electrons move **up** the potential gradient. In solids, and in a vacuum, so far as we know, **electrons are the only electrical entities that move.** Due, however, to the original single-fluid (positive) theory of Franklin, it has been the custom to regard

the direction of the current as taking place from points of high potential to points of low potential, *i.e.*, **down** the potential gradient. This is not in conformity with known facts. Unfortunately the older convention is so thoroughly established in the literature of the subject, particularly in the field of electrical engineering, that it will probably persist until one or more of the learned societies act officially to abandon the custom. If, under all circumstances, the direction of the electron flow was to be considered as the direction of the current, it would be necessary to change several of the presently used standard relations having to do with electric and magnetic fields. Therefore, when it seems advisable to deal with the current as a flow of positive entities, we shall follow the custom of referring to the conventional current simply as the **current;** if and when the electron flow is meant, the term **electron current,** or electron movement, will be used. In the diagrams to follow a simple arrow will imply that the conventional current (transfer from positive to negative) is indicated; when the electron current is meant, an arrow followed by a small letter **e** will be used.[1]

69. Ohm's Law. The relation most frequently employed in dealing with the electric current was disclosed by Dr. F. S. Ohm, a German mathematician-physicist, in a paper published in 1826. Ohm found that in solids the magnitude of the current in such a conductor is strictly proportional to the potential difference which is maintained at its terminals, *i.e.*, $I \propto V$, where V is the potential difference involved. The case may be expressed in another way by saying that the ratio of the potential difference to the current is a constant. Algebraically this statement would take the form

$$\frac{V}{I} = \text{a constant.}$$

This constant is known as the resistance of the conductor, and is commonly designated by the symbol R. The above expression may accordingly be written in any one of several useful forms, viz,

$$\frac{V}{I} = R; \qquad \frac{V}{R} = I; \qquad V = RI. \tag{91}$$

There are certain limitations as to the use of Ohm's law. The law is applicable only to those cases in which the applied potential is constant in magnitude and direction. Further, the law is not applicable to all classes of conducting mediums, as we shall see later. Notwithstanding these definite limitations, **Eq. (91) expresses one of the most useful relations in the whole realm of electrical science.**

[1] This convention conforms to the practice now being followed by several authors.

70. Resistance. The resistance term in Ohm's law also calls for detailed examination. **Resistance may be defined as that property of a conductor by virtue of which the energy of moving electrons is converted into heat.** The reciprocal of resistance ($1/R$) is referred to as **conductance.**

In the course of his research Ohm found that the resistance of a given conductor is a function of its length, and of its cross-sectional area. Mathematically his findings may be expressed thus,

$$R = \rho \frac{l}{a}, \tag{92}$$

where l represents the length, a the cross-sectional area, and ρ a proportionality constant known as **specific resistance,** or **resistivity.** The magnitude of ρ depends upon the nature of the material constituting the conductor. If we make l and a unity we have a definition of ρ which would be to the effect that specific resistance is the resistance of a conductor whose length is one centimeter and whose cross section is one square centimeter. In engineering work the foot is taken as the unit of length and the circular mil[1] as the unit of cross-sectional area. Such a portion of a conductor is referred to as a mil-foot. Thus ρ may be expressed in either of two ways, one value being based on the resistance of a centimeter cube of the material, and another involving the resistance of a piece of the material 1 ft long and 1/1,000 in. in diameter.

In Ohm's law, if V is expressed in statvolts and I in statamperes, R will also be in cgs electrostatic units, viz., statohms. In the expression $V/I = R$, if we change V to volts and I to amperes, the left side of this equation would be $\frac{V \times 300}{I/3 \times 10^9}$, or $\frac{V}{I} \times 9 \times 10^{11}$. To preserve equality in the expression for Ohm's law, it will therefore be necessary to multiply R by 9×10^{11} also. This means that we would have a new unit of resistance whose value would be $1/9 \times 10^{-11}$ of the electrostatic unit. This unit is known as the ohm; and **a conductor is said to have a resistance of one ohm when a potential difference of one volt applied at its terminals will cause a current of one ampere to flow.** By an international agreement arrived at in 1908 there was established what is known as the international ohm. This standard unit is the resistance of a column of mercury of uniform cross section at 0°C and having a length of 106.300 cm and a mass of 14.4521 gm. In the table which follows will be found

[1] A mil is one-thousandth of an inch. A circle whose diameter is 1 mil has an area of 1 cir mil. Because of the fact that the area of any circle, when expressed in circular mils, is numerically equal to the square of the diameter of the circle in mils, this method of expressing areas is found to be convenient in practical engineering.

the values of ρ for a number of materials commonly used as electrical conductors. Note that ρ is given in both systems of units. Since the resistance of a centimeter cube of most conducting material is very small, ρ, in the metric system, is expressed in **microhms** rather than in ohms.

RESISTIVITY AND TEMPERATURE COEFFICIENT[1]

Material	Composition or condition	Resistivity		Temperature coefficient, ohms per ohm per degree C
		Microhm per cm cube	Ohms per mil-ft	
Advance	Cu-Ni	48.8	294	+0.000020
Aluminum	Wire annealed	2.83	17.02	+0.0039
Copper	Wire annealed	1.724	10.4	+0.0040
Carbon	0°C	3/500 00	......	−0.0003
Constantan	Copper (60%) Nickel (40%)	49.00	300	−0.000005
Climax	Ni-Steel	87.00	525	+0.00054
German silver	Cu-Ni(30%)Zn	33.00	290	+0.0004
Gold	Pure	2.44	12.42	+0.0037
Iron	Pure-ann.	10.00	59.9	+0.0045
Iala	Cu-Ni	49.0	283	−0.000005
Manganin	Cu-Mn-Ni	44.0	265	+0.000006
Mercury	at 0°C	94.07	....	+0.00072
Nickel	Ni	7.8	64.3	+0.0062
Nichrome	Ni-Cr	109.0	660	+0.00019
Platinum	Pt at 0°C	10 96	66.0	+0.00366
Silver	Ag at 0°C	1.5–1.7	9–10.2	+0.0038
Tungsten	W	5.51	33.2	+0.005
Rubber	....	2×10^9		
Impregnated paper	....	5×10^8		
Varnished cambric	....	2×10^8		
Glass	....	5×10^{10}		
Fused quartz	....	1×10^{12}		

[1] Values are for 20°C unless otherwise specified.

In dealing with resistivity, it is to be borne in mind that there are a number of factors that tend to modify the resistance of a conductor. One of these is temperature. For moderate differences in temperature, the relation between resistance and temperature is approximately linear. This relation is given by the expression

$$R_t - R_0 = \alpha R_0 t,$$

in which R_t is the resistance of the conductor at temperature t°C; R_0 its resistance at 0°C; and α the proportionality constant, which in this case is the temperature coefficient of resistance. A definition of α may be derived by changing the above equation into the form

$$\alpha = \frac{R_t - R_0}{R_0 t}. \tag{93}$$

An examination of the above equation shows that the temperature coefficient of resistance may be defined as the fractional change in resistance that a conductor undergoes per degree change in temperature. Equation (93) indicates that α would be expressed in ohms per ohm per degree of temperature.

It is to be noted that most of the pure metals commonly used as conductors show relatively **high** temperature coefficient values. On the other hand, most of the resistance alloys exhibit relatively **low** temperature coefficients. In fact, in one or two cases the coefficient is vanishingly small, **constantan** and **manganin** being notable examples. Thus there are available several alloys that have high resistivity and low temperature reactions. These materials are widely used as material from which to make wire that is to be utilized in the designing of units that are to serve as standards of resistance. Manganin is most commonly used for this purpose. A few substances have negative temperature coefficients, the most notable example being carbon. Most insulating materials show a negative coefficient.

In order to determine experimentally the value of α for any given material it will obviously be necessary to determine the resistance at two temperatures, one of which is zero. By any one of several methods, to be described in the next chapter, one may readily measure the resistance at any convenient working temperature, and also at zero degrees. The necessity of measuring the resistance at zero degrees may be avoided by one of two procedures. Equation (93) may be written as

$$R_1 = R_0(1 + \alpha t_1),$$

where R_1 is the resistance of the conductor when it is at temperature t_1. If now we find its resistance at some other temperature, say t_2, we may write

$$R_2 = R_0(1 + \alpha t_2).$$

Eliminating R_0 between the last two equations, we get

$$\alpha = \frac{R_2 - R_1}{R_1 t_2 - R_2 t_1}. \tag{94}$$

Thus by measuring the resistance at two known temperatures one can

readily determine the average value of α for that particular temperature range.

Another method involves the taking of several resistance-temperature readings and plotting the resulting data, as indicated in Fig. 66. The graph shown was set up from data taken on a sample of pure nickel

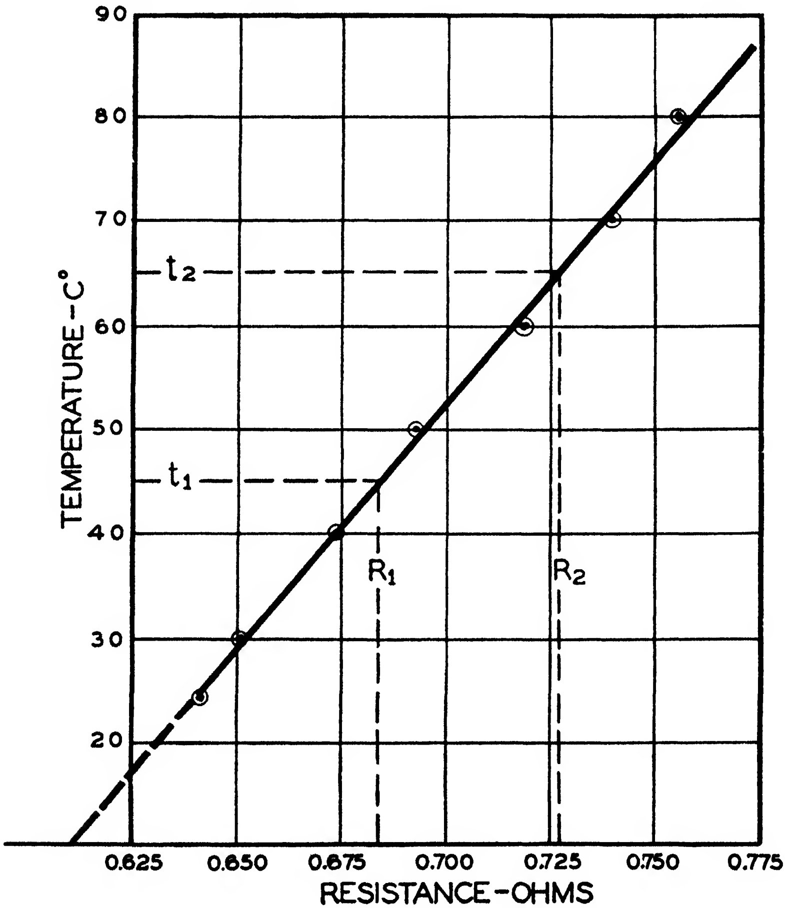


FIG. 66.—Graph showing the effect of temperature on the resistance of a conductor. Specimen tested was pure nickel wire.

wire. By extending the curve until it cuts the resistance axis one can easily determine the R_0 value and thus compute α by the use of Eq. (93). As a check, one may evaluate α by determining the slope of the curve.

For wider ranges of temperature the relation

$$R_t = R_0(1 + \alpha t + \beta t^2) \tag{95}$$

will give more accurate results, in which β is another constant that must be determined.

The temperature-resistance laws given above do not hold for extreme temperature values. For instance, it was found by Onnes that at the temperature of liquid helium ($-267°C$) the resistance of pure metals is extremely low; in fact, much lower than might be expected from the foregoing relations. It has been found, for example, that an electric current once set up in a ring of pure lead which is maintained at the temperature of liquid helium will continue to flow for a length of time measured in hours, and several days elapse before it drops to half its original value. It would thus appear that at temperatures in the region

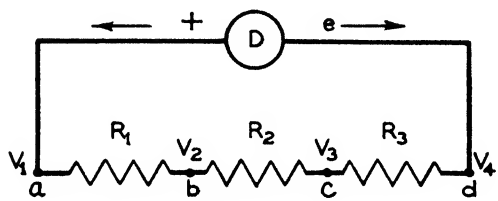


FIG. 67.—Resistances in series.

of absolute zero the resistance of at least some conductors approaches zero.

The presence of small amounts of chemical impurities, however, tends to **increase** the resistivity of most metals. For example, the presence of 1 per cent of carbon will increase the resistivity of copper as much as 20 per cent. In general, alloys have a higher resistivity value than is shown by the individual constituents which go to make up the alloy. An examination of any one of several cases listed in the resistivity table shown on page 116 will bear out the foregoing statement. Take, for instance, the case of the resistance alloy known as constantan. This alloy consists of 60 per cent copper and 40 per cent nickel; ρ for Cu is 1.724 microhm-cm and for nickel it has a value of 7.8. The alloy has a resistivity of 49.

An alloy extensively used in heating devices is known as nichrome. In addition to having a high resistivity value, its temperature coefficient is fairly low and, it will withstand high temperatures without changing chemically. A few alloys show a large and an **abrupt** change in resistance at certain temperatures. Such materials are useful in connection with the design of what are called **"ballast" resistors**—devices that serve to limit automatically the flow of current.

71. Resistances in Series and Parallel. A circuit may be made up of several resistance units connected in series as shown in Fig. 67. Suppose

the electronic current to be flowing in the circuit in the direction indicated by the arrow, *i.e.*, from d toward a.[1] The point a is therefore considered to be at a higher potential than d, and accordingly there will be a fall in potential from a to d. The magnitude of the current will be the same at all points in the circuit. Bearing in mind the fact that Ohm's law is applicable not only to the circuit as a whole, but also to each component, we may write

$$V_1 - V_2 = R_1 I,$$
$$V_2 - V_3 = R_2 I,$$
$$V_3 - V_4 = R_3 I.$$

The total potential drop between a and d will obviously be given by the sum of the above expressions, or

$$V_1 - V_4 = (R_1 + R_2 + R_3)I.$$

This is equivalent to

$$\frac{V_1 - V_4}{I} = R_1 + R_2 + R_3 = R, \tag{96}$$

where R is the total resistance between a and d, and $V_1 - V_4$ is the total drop (fall in potential) over the combined resistances. It may therefore be said that **the total resistance of several resistances in series is equal to the sum of the individual resistances.**

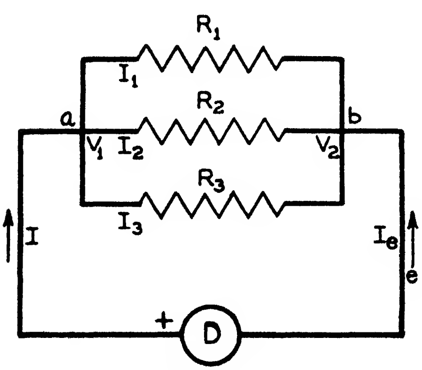


FIG. 68.—Resistances in parallel.

A second arrangement of resistance components is sketched in Fig. 68, the three resistances being taken as representative. Such a network is spoken of as a parallel, or multiple, arrangement of resistances. If an electronic current I flows in the main circuit, it will, upon reaching b, divide into three parts as shown. (The conventional current will divide at a.) Let R be the combined resistance between a and b. The potential of the points a and b are V_1 and V_2, respectively, and these values are fixed. **The fall in potential over each branch will therefore be the same.** Again applying Ohm's law we have

[1] The traditional current direction would be the reverse of this, *i.e.*, from a toward d. But no electrical entities actually move in that direction.

$$I_1 = \frac{V_1 - V_2}{R_1},$$

$$I_2 = \frac{V_1 - V_2}{R_2},$$

$$I_3 = \frac{V_1 - V_2}{R_3},$$

etc., for any number of resistors. Now

$$I = \frac{V_1 - V_2}{R},$$

and since the sum of the currents in the several branches must equal the current in the supply circuit, we may write

$$\frac{V_1 - V_2}{R} = \frac{V_1 - V_2}{R_1} + \frac{V_1 - V_2}{R_2} + \frac{V_1 - V_2}{R_3} + \cdots + \frac{V_1 - V_2}{R_n},$$

or

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n}. \tag{97}$$

The foregoing relation shows that the combined resistance of several conductors in parallel is less than the resistance of any one of the individual components. It also shows that the total conductance equals the sum of the conductances of the individual branches. For a network of three branches, as illustrated above, Eq. (97) reduces to the more convenient form

$$R = \frac{R_1R_2R_3}{R_1R_2 + R_2R_3 + R_1R_3}. \tag{98}$$

In the case of two branches this becomes

$$R = \frac{R_1R_2}{R_1 + R_2} \tag{99}$$

Again referring to Fig. 68, and considering, for our present purposes, that there are only two branches instead of three as shown, we have a case frequently encountered in practice. For instance, a current by-pass or **shunt** may be connected about the internal winding of a current-indicating instrument such as a galvanometer or ammeter. The question will arise as to the value of the current in each branch.

As outlined above

$$I_1 = \frac{V_1 - V_2}{R_1}$$

$$I_2 = \frac{V_1 - V_2}{R_2},$$

or

$$I_1R_1 = V_1 - V_2 = I_2R_2.$$

Therefore

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}, \tag{100}$$

which shows that the current will divide inversely as the resistance of the individual branches. Using this equation, and the relation $I = I_1 + I_2$, it thus becomes possible in any given case to compute the current in either the main or shunt circuit.

Problem. Suppose we have a galvanometer whose resistance is 80 ohms, and it is desired to arrange a shunt in parallel with the instrument winding so that only $\frac{1}{100}$ of the total current shall pass through the galvanometer. What must be the value of the shunt resistor?

Solution. Adapting Eq. (99) to the case in hand, we have

$$\frac{I_g}{I_s} = \frac{R_s}{R_g},$$

where the subscripts g and s indicate galvanometer and shunt respectively. Since the galvanometer winding is to carry $\frac{1}{100}$ of the total current the ratio of the galvanometer current to the current in the shunt (I_g/I_s) will be $\frac{1}{99}$ and hence we have

$$\frac{1}{99} = \frac{R_s}{80},$$

which leads to

$$R_s = 0.808 \text{ ohm}$$

as the resistance of the shunt.

72. Kirchhoff's Laws. Two useful corollaries, known as **Kirchhoff's laws,** or rules, may be deduced from Ohm's law. The first of these generalizations is to the effect that in any given network where several conductors meet at a common point the sum of the currents flowing toward the point is equal to the sum of the currents flowing away from the point. In other words, **the algebraic sum of the currents entering and leaving a common point is zero.** It is the convention to use the positive algebraic sign for the currents flowing toward the point and the negative sign for those which flow away. In general terms one might write

$$\Sigma(\pm I_1 \pm I_2 \pm I_3 \pm \cdot \cdot \cdot \pm I_n) = 0. \tag{101}$$

In the case illustrated by Fig. 69 we may write

$$I_1 - I_2 + I_3 - I_4 + I_5 = 0.$$

This law follows from the fact that, in the case of a constant current in a circuit of negligible capacitance, the electrons do not accumulate at any one point.

Kirchhoff's second law is a statement of the important fact that, in any closed circuit, the algebraic sum of the several IR products is equal to the algebraic sum of the emf's acting in that particular path. Mathematically the case could be stated thus

$$\Sigma IR + \Sigma \text{emf} = 0. \tag{102}$$

The logic of the second law becomes apparent if one starts at some given point in a circuit loop and makes an electrical journey completely around the loop. On returning to the starting point one must be at the same potential level as at the start. In other words, the sources of emf encountered in the journey must yield a total emf **equal to the total potential drop due to the resistances through which we pass.** In making such an electrical journey around a circuit due regard must be given to the **sense** of the several IR and emf terms. In passing through a source of emf from the negative to the positive terminal the potential increases, and therefore the emf term should carry a + sign. Conversely, when passing through an emf source from the positive terminal to the negative connection, there is a fall in potential and the emf term should be given a − sign.

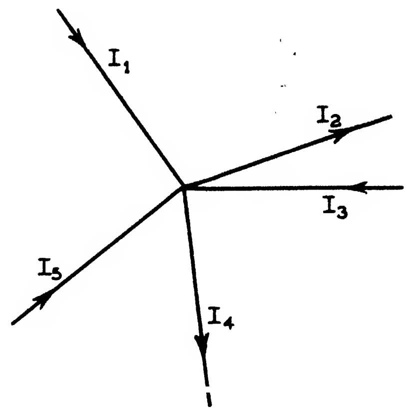


FIG. 69.—Illustrating Kirchhoff's first law.

When passing through a resistance **of any type,** in the direction of the conventional current, an IR drop will carry a **negative** sign. When going through a resistance in the opposite direction, *i.e.*, against the conventional current, the IR drop is given a **positive** sign. If we follow the electronic current these signs will be reversed.

In order to apply Kirchhoff's law to the solution of problems involving networks it is important to keep in mind two definite considerations:

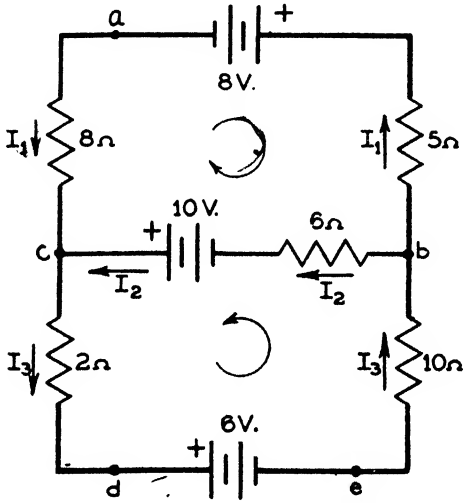
1. The first law must be applied to enough junctions so that each current appears at least once.
2. The second law must be applied enough times to include each path at least once.

If these precautions are observed the total number of resulting equations will equal the total number of unknown factors.

If, in first analyzing a given problem, the direction of some particular

current is not known we may **assume** it to flow in a definite direction. If the numerical result shows a **positive** algebraic sign our assumption was correct; if a **negative** sign, it means that the actual direction of the current is the reverse of that which we assumed. Incorrect assumptions as to current directions **will not alter** the numerical results.

Kirchhoff's laws are among the most useful generalizations available in electrical study. By the aid of Ohm's law and Kirchhoff's laws it is possible to solve any problem involving a d-c network, no matter how complicated the network may be. Indeed it is also possible to apply Kirchhoff's laws to a-c problems, as will be pointed out later. The utility of Kirchhoff's laws may be illustrated by applying them to a specific case, as shown in the following numerical problem.



Problem. Find the magnitude and direction of the currents in the network sketched above. The internal resistances of the sources of emf (in this case, batteries) may be neglected.

Solution. The direction of the several currents not being known, it will be assumed that they are as indicated by the straight arrows. We will first apply Kirchhoff's second law to the loop *abca*, beginning at, say, *a*. This procedure will give

$$8 + 5I_1 - 6I_2 + 10 + 8I_1 = 0,$$

which reduces to

$$18 + 13I_1 - 6I_2 = 0. \qquad \text{(i)}$$

Next proceeding around the loop *bcdeb*, we have

$$-6I_2 + 10 - 2I_3 - 6 - 10I_3 = 0.$$

Reducing,

$$4 - 6I_2 - 12I_3 = 0. \qquad \text{(ii)}$$

We have three unknowns; hence it will be necessary to have at least three equations in order to complete the solution. To secure another equation we may apply Kirchhoff's first law to some point, say, c. This gives us

$$I_1 + I_2 - I_3 = 0 \qquad \text{(iii)}$$

Combining (i), (ii), and (iii), it is found that

$$\begin{aligned} I_1 &= -0.98 \text{ amp} \\ I_2 &= 0.875 \text{ amp} \\ I_3 &= -0.105 \text{ amp}. \end{aligned}$$

The negative sign before the current value for I_1 and I_3 indicates that these currents actually are in directions opposite to those that were assumed.

PROBLEMS

1. An incandescent lamp taking 0.5 amp is turned on for 5 min. How many electrons passed through the filament of the lamp during that period?

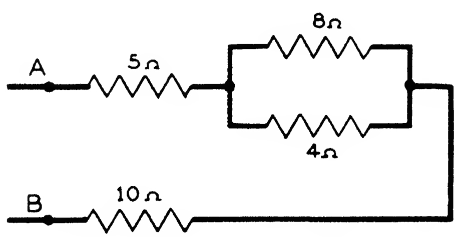
2. No. 30 copper wire has an approximate area of 10 cir mils. What is the resistance of 1,000 ft of such wire at 20°C?

3. No. 4 wire has a diameter of approximately 0.52 cm. What will be the resistance of 100 m of such wire at 20°C?

4. What would be the resistance of the wire indicated in Prob. 2 at a temperature of 60°C?

5. The resistance element of a platinum thermometer at 0°C is 25 ohms. When immersed in a certain solution the resistance is found to be 30 ohms. What is the temperature of the solution?

6. Assuming the resistance values shown in the sketch, find the resistance between the points A and B. If a potential difference of 50 volts is maintained

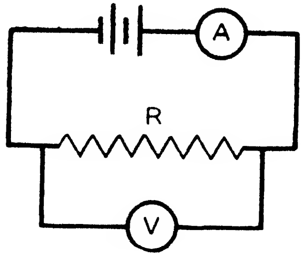


between A and B what will be the magnitude of the current? What will be the fall in potential across the 8-ohm resistor? Across the 4-ohm unit?

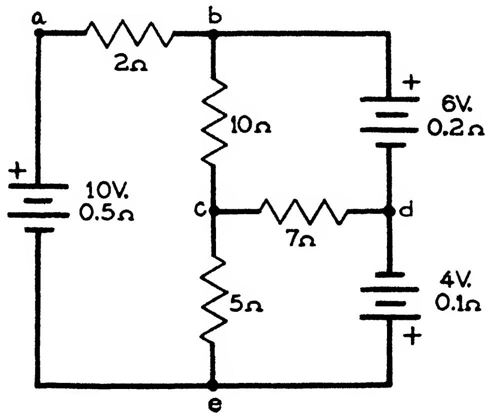
7. The coil of a certain galvanometer has a resistance of 120 ohms. It is desired to arrange a shunt of such a resistance value that only 1/1,000 of the current to be measured will pass through the galvanometer winding. What should be the resistance of the shunt?

8. A battery cell shows an emf of 2.2 volts on open circuit. When its terminals are connected by a conductor whose resistance is 10 ohms the current is 0.2 amp. What is the terminal voltage of the cell, and what is its internal resistance?

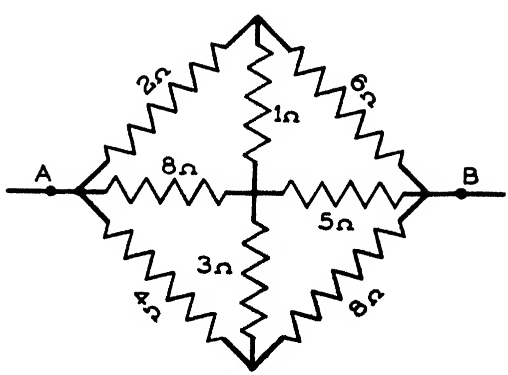
9. In the circuit shown in the accompanying diagram the ammeter reads 25 ma and the voltmeter 0.25 volt. What is the value of the resistance R?



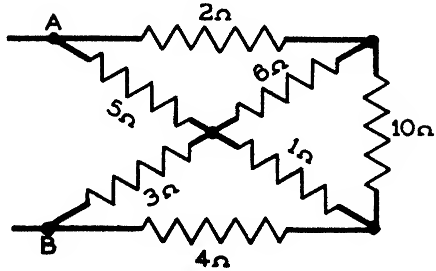
10. Assume a network as shown in the sketch. Determine: (*a*) The magnitude and direction of all currents; (*b*) the drop over *cd*, *ab*, and the 10-volt battery.



11. The sketch below shows a network the components of which have the resistances indicated. Find the resistance between the points A and B.



12. In the network shown, what will be the current in the 10-ohm resistor if a difference of potential of 10 volts is applied between the points A and B?



CHAPTER IX

FUNDAMENTAL DIRECT-CURRENT MEASUREMENTS

73. The Wheatstone Bridge. The determination of resistance values is probably the most important electrical measurement that one is called upon to make, and with suitable equipment it is possible to make measurements of this character to an accuracy better than $\frac{1}{100}$ of 1 per cent.

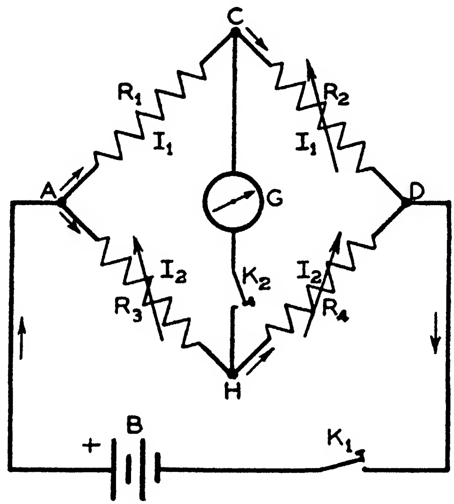


FIG. 70.—Wheatstone-bridge network.

The most widely used method of measuring resistance was devised by Wheatstone and consists of a network made up, essentially, of four conductors as shown in Fig. 70. In the diagram G represents a current-detecting device, such as a galvanometer; B is a source of emf. Keys K_1 and K_2 are provided for closing the battery and galvanometer circuits as shown.

Suppose a difference of potential is applied between the points A and D of the network. Current (conventional) will flow along the paths ACD and AHD. For a given potential difference between A and D the magnitude of the current I_1 which will flow through the branch ACD will depend upon the resistances R_1 and R_2. Likewise the current I_2 will depend upon the value of R_3 and R_4. The potential drop over AC will be given by I_1R_1, and likewise for AH by I_2R_3. Now if and when the

several resistances comprising the network are so adjusted in value that no current passes through the galvanometer G it follows that the points C and H are at the same potential, or

$$I_1R_1 = I_2R_3.$$

When this condition obtains it also follows that

$$I_1R_2 = I_2R_4.$$

Dividing one of these equations by the other and simplifying, we have

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}. \tag{103}$$

If, then, three of these resistances are known, the fourth can be readily determined.

In practice there are two common forms of the Wheatstone bridge. One, known as the slide-wire type, consists of a resistance wire of uniform cross section, corresponding to R_3 and R_4, on which rests a sliding contact corresponding to H in Fig. 70. By this arrangement it is possible to vary the ratio of R_3 to R_4. Since the resistance constituting R_3 and R_4 is uniform in cross section, and since resistance is proportional to length, length may be substituted for resistance in Eq. (103), thus giving

$$\frac{R_1}{R_2} = \frac{L_1}{L_2} \tag{104}$$

where L_1 and L_2 correspond to R_3 and R_4 respectively. The resistance R_1 may be the unknown and R_2 a standard resistance coil, or set of such coils,

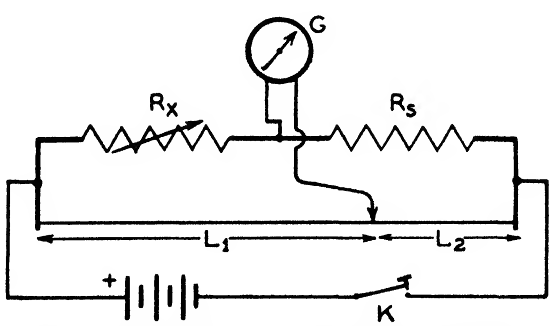


FIG. 71.—Slide-wire form of Wheatstone bridge.

usually in the form of a so-called "resistance box." Figure 71 is a diagrammatic sketch of an assembly of this type. The resistance wire is commonly a meter in length and is stretched above a meter stick. Heavy copper bars of negligible resistance serve to connect the ends of the slide

wire to the openings provided to receive the standard and unknown resistances, R_s and R_x. In practice, provision is made for conveniently interchanging the standard and unknown resistances, thus eliminating, in part, the effect of contact resistance between these resistances and the connecting bars.

A second form of the Wheatstone bridge finds wide application in engineering practice. Because it was originally made for use in the English[1] postal service, it is known as the "post office" box or bridge. In this form of the bridge the resistances R_3 and R_4 consist of a set of "ratio coils" of fixed values. By means of contact plugs or a wiping switch, ratios from $1/10$–$1/1000$ and from 1,000 to 1 may be cut into the bridge circuit. The resistance R_2 consists of a set of accurately made resistance coils, assembled in a so-called "decade" arrangement whereby it is possible to secure resistance values ranging from 1 to 9,999 ohms or, in some cases, higher. Any one coil or a combination of coils may be introduced into the bridge network by means of contact plugs or a rotating switch. It thus becomes possible to measure resistance over a wide range of values. In the portable form of the post office bridge, or **testing set,** as it is sometimes called, a portable galvanometer and a dry battery are housed in the same case as the resistance coils. A commercial unit of this type is shown in Fig. 72.

FIG. 72.—Portable Wheatstone bridge, or test set. (*Leeds & Northrup Co.*)

74. Carey-Foster Bridge. Another and somewhat more accurate method of measuring resistances, particularly slight changes in resistance, was devised by Professor Carey-Foster. This arrangement consists of a network made up essentially of six resistances, one of which is a standard resistance of high accuracy. From the following description it will be apparent that the Carey-Foster method makes it possible to determine **the difference between two resistances,** one of which is of standard value, to a very high degree of accuracy.

The unique feature of the Carey-Foster modification of the more simple Wheatstone bridge consists in the interchange of the standard resistance and the resistance under test in such a manner as to eliminate the resistance of the connecting bars.

In Fig. 73, S_1 and S_2 are the two resistances being compared; R_1 and

[1] In England the telegraph and telephone service is operated by the government post-office department.

R_2 are two nearly equal auxiliary resistances. The exact value of these auxiliary resistances need not be known, as they disappear from our final relation. Let r_1 and r_2 be the resistances of the connecting strips MN and $M'N'$, respectively, and ρ be the resistance of unit length of the bridge wire. NN' is a uniform resistance wire similar to that incorporated in the original Wheatstone bridge. It is to be noted that the resistances representing S_1, r_1, and a constitute one arm of the bridge, while S_2, r_2, and b together form the other corresponding arm. In prac-

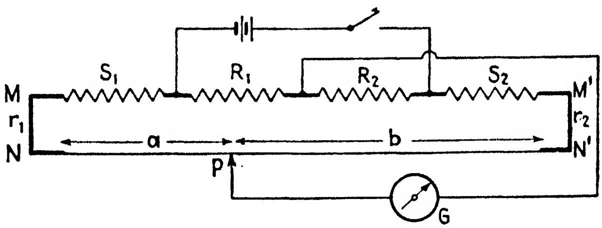


FIG. 73.—Carey-Foster bridge network.

tice a standard resistance unit, say S_1, is selected which has a value very nearly that of the unknown S_2. The galvanometer contact p is then adjusted until zero deflection obtains, and the readings a and b are noted. By means of a convenient switching arrangement the standard and unknown resistances are then interchanged and another balance secured. In general p will now show a different setting, and the resistance of that portion of the bridge wire between the first and second settings will equal the difference between the values of the standard and unknown resistances. For the first setting we may write

$$\frac{R_1}{R_2} = \frac{S_1 + r_1 + \rho a_1}{S_2 + r_2 + \rho b_1}, \qquad \text{(i)}$$

where a_1 and b_1 represent the lengths of the bridge wire to the left and right of the contact point p.

When S_1 and S_2 are interchanged, if a_2 and b_2 be the readings on the bridge wire for the new balance, we have

$$\frac{R_1}{R_2} = \frac{S_2 + r_1 + \rho a_2}{S_1 + r_2 + \rho b_2}. \qquad \text{(ii)}$$

It is desired to eliminate the factors R_1, R_2, r_1, and r_2 from our equations. To do this we may add unity to both sides of (i) and (ii). This gives

$$\frac{R_1 + R_2}{R_2} = \frac{S_1 + r_1 + \rho a_1 + (S_2 + r_2 + \rho b_1)}{(S_2 + r_2 + \rho b_1)} \qquad \text{(iii)}$$

and

$$\frac{R_1 + R_2}{R_2} = \frac{S_2 + r_1 + \rho a_2 + (S_1 + r_2 + \rho b_2)}{(S_1 + r_2 + \rho b_2)}. \tag{iv}$$

It is obvious that the right sides of the above two equations constitute an identity.

Further, since changing the position of the galvanometer contact p does not alter the total resistance of the bridge, it follows that

$$a_1 + b_1 = a_2 + b_2.$$

By rearranging terms, and utilizing the above relation, it will be evident that the numerators of the right-hand members of (iii) and (iv) are equal, and hence

$$S_1 + r_2 + \rho b_2 = S_2 + r_2 + \rho b_1$$
$$S_1 - S_2 = \rho(b_1 - b_2) = \rho(a_2 - a_1). \tag{105}$$

Thus we see, as previously indicated, that the difference between the resistance of the two coils S_1 and S_2 is equal to the resistance of that part of the bridge wire between the points at which the contact p is set to secure the balances for the two positions of the resistances.

It will be observed that our final relation above involves ρ, the resistance of unit length of the bridge wire. This constant must of course be known or determined in advance. In practice a Carey-Foster bridge has several wires of different resistance. The method to be followed in determining the value of ρ for each wire will depend to some extent upon the resistance of the particular wire being calibrated.

The last equation above may be written in the form

$$\rho = \frac{S_1 - S_2}{a_2 - a_1}. \tag{106}$$

Hence if the difference in the resistance of the two coils S_1 and S_2 is known, ρ can be found by determining two successive balances. If the total resistance of a given bridge wire be of the order of $\frac{1}{10}$ ohm, S_1 may be one 1-ohm standard and S_2 may consist of a 10- and a 1-ohm standard unit in parallel. This would give a value of $\frac{10}{11}$ ohm for S_2; hence $S_1 - S_2$ would be $\frac{1}{11}$ or 0.09091 ohm.

If the resistance of a bridge wire is of the order of $\frac{1}{100}$ ohm, a 1-ohm coil may be used on one side and a 1-ohm and 100-ohm unit in parallel on the other. This would give for $S_1 - S_2$ a value of 0.009901 ohm.

If the entire bridge wire be considerably in excess of 1 ohm, then ρ may be found by the aid of a single standard resistance unit and a heavy copper bar or link, the resistance of the latter being negligible; in which case

$$\rho = \frac{1}{a_2 - a_1}.$$

Standard resistances are made of manganin wire, the coils being carefully aged after winding. The coil itself is inclosed in a metal case containing oil. The value of such standards are usually certified by the maker. Heavy copper leads are provided for making connections to the

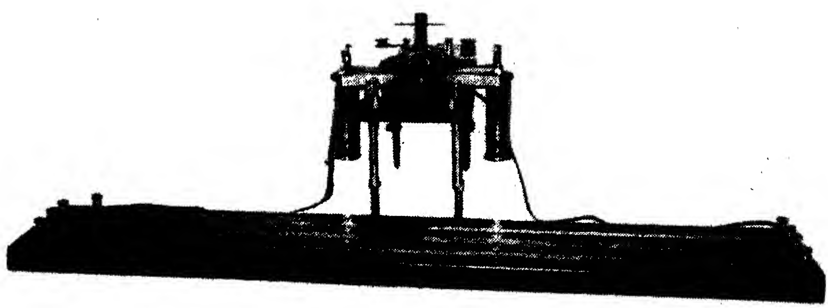


FIG. 74*a*.—Laboratory form of Carey-Foster bridge.

bridge. Figure 74 shows one form of a Carey-Foster bridge and a typical standard resistance unit.

One important use of the Carey-Foster network is in the determination of the temperature coefficient of a conductor [Sec. 70, Eqs. (93) and (94)]. It is also used to check the resistance values of the coils in resistance boxes.

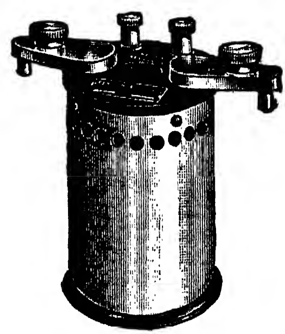


FIG. 74*b*.—Standard resistance for use with Carey-Foster bridge. (*Leeds & Northrup Co.*)

75. Kelvin Double Bridge. In the measurement of resistances of the order of 0.0001 ohm the resistance of the connecting leads and contact resistance may be comparable in magnitude with the resistance of the sample under test. These possible sources of error are eliminated by means of a resistance network devised by Lord Kelvin. In addition, the low sensitivity of both the Wheatstone and Carey-Foster bridges when measuring low resistances is overcome by the Kelvin arrangement. Instead of having one set of ratio coils as in the two types of bridges already described, the Kelvin network employs **two** sets of ratio coils, hence the term "double." A diagrammatic sketch of the network is shown in Fig. 75.

In the diagram X represents the bar of metal whose resistance is to be measured, and S is the resistance bar which has been very carefully

calibrated in fractions of an ohm. One end C of the standard and one end B of the unknown are connected by a heavy low-resistance bar J, and **also by a pair of ratio coils c and d.** These auxiliary resistors (c and d) constitute the **only, and the important, difference** between the Kelvin and Wheatstone networks. The introduction of these two auxiliary resistors makes it possible to eliminate the resistance of this connecting strip and associated contact resistances from the final consideration. a and b also constitute another set of ratio coils. These figure in the final

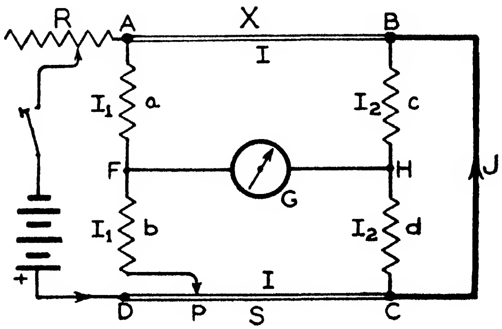


FIG. 75.—Kelvin double-bridge network.

relations, as we shall see. The resistance of all four of the ratio coils is high compared with that of the standard and the sample. One end of one of the ratio coils is arranged as a traveling contact p on the standard resistance bar. A current of several amperes is supplied by a battery, the current magnitude being controlled by the rheostat R. The first step in adjusting the Kelvin bridge consists in selecting a ratio value a/b adapted to the problem in hand. Greatest accuracy will be secured by making a/b of such a magnitude that as much of the standard resistance as possible is included in the circuit when balance is secured. **The ratios a/b and c/d should be equal.** The practical construction of the bridge is such that this equality is readily attained.

When the galvanometer G shows zero current the points F and H are at the same potential. This condition of balance is brought about by adjustment of the contact p, thus varying the portion of the standard S included in the circuit.

When balance is attained a current I_1 flows through a and b; and likewise a current I_2 flows through c and d. Also the current flowing through X and S has the same value, say I. Applying Ohm's law we have

$$aI_1 = XI + cI_2 \quad \text{(i)}$$
$$bI_1 = SI + dI_2. \quad \text{(ii)}$$

Rewriting (i) and (ii), we have

$$XI = aI_1 - cI_2 \tag{iii}$$
$$SI = bI_1 - dI_2. \tag{iv}$$

The expressions embodied in (iii) and (iv) may be changed to the forms

$$XI = a\left(I_1 - \frac{c}{a}I_2\right) \tag{v}$$
$$SI = b\left(I_1 - \frac{d}{b}I_2\right). \tag{vi}$$

Initially it was noted that $a/b = c/d$; hence $d/b = c/a$. Dividing (v) by (vi) and utilizing the equality just cited, we have

$$\frac{X}{S} = \frac{a}{b}. \tag{107}$$

Thus we see that the resistance of the connecting bar J, together with that of the connecting strips and contacts, is eliminated. In fact, the method is seen to be an arrangement for comparing potentials. To

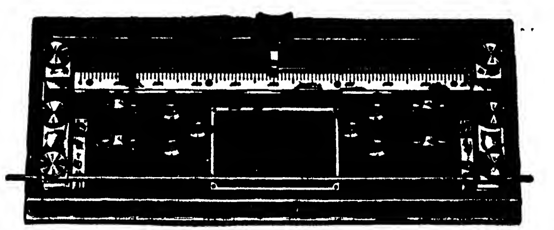


FIG. 76.—Simple form of Kelvin bridge. (*Leeds & Northrup Co.*)

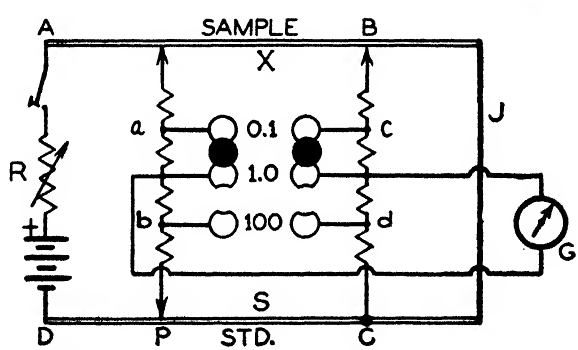


FIG. 77.—Schematic wiring diagram of connections for the Kelvin bridge shown in Fig. 76.

increase the drop over the sample and over the standard bar, and thus increase the sensitiveness, the current may be made large, the only

limitation in this respect being that these components shall not be appreciably heated. Bridges of this type are commonly so arranged that a/b ratios of 1:1, 1:10, and 1:100 may be had. The Kelvin bridge is particularly useful in determining the specific resistance of various metals. Figure 76 shows a simple form of this bridge and Fig. 77 gives a schematic

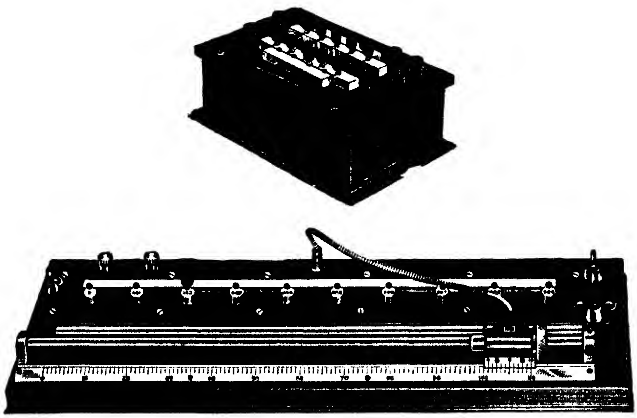


FIG. 78.—Commercial form of Kelvin bridge. *Above*, ratio coil box; *below*, variable standard resistance. (*Leeds & Northrup Co.*)

wiring diagram of the unit. The lettering corresponds with that of Fig. 75. Figure 78 is an illustration of one commercial form of the Kelvin bridge. Such units are widely used in the testing of heavy conducting wire and bars, such as trolley wire and bus-bar stock. When specially arranged for such work they are sometimes referred to as **conductivity bridges.**

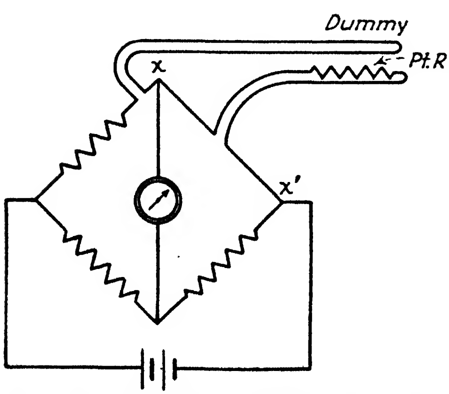


FIG. 79.—Resistance-thermometer network.

76. Resistance Thermometry. Reference to the table on page 116 will disclose the fact that certain substances exhibit a relatively high temperature coefficient and in one or two cases the change in resistance follows any change in temperature very accurately. This is particularly true of platinum and nickel, especially the former. Advantage is taken of this property in the construction of what are known as resistance thermometers. Figure 79 shows an electrical network designed for use in making temperature measure-

ments. In order to indicate the temperature of the sensitive arm of the bridge, one of two plans may be followed. A variable resistance may form the conjugate arm of the bridge and be used to restore the balance as the resistance of the arm xx' changes. This variable resistance may be calibrated in degrees of temperature. The device thus becomes direct reading.

Another possibility consists in having the resistance of the three nonsensitive arms of the bridge of fixed value and employing a deflection method. As the resistance of arm xx' changes, the galvanometer will show a deflection as a result of the unbalance of the bridge. The galvanometer can thus be calibrated to read directly in degrees of temperature.

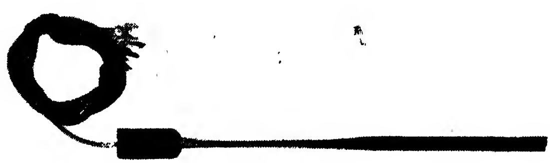


FIG. 80.—Resistance thermometer—calorimeter type. (*Leeds & Northrup Co.*)

Dummy leads of the same length and material as those which serve to connect the platinum resistance are inserted in the conjugate arm and housed in the same casing. This is to compensate for any change in the resistance of the thermometer leads.

The housing of the resistance wire takes on various forms, depending upon the use to which the equipment is to be put. For instance, when the temperature of molten metal is to be observed the resistance element is inclosed in a porcelain or pyrex jacket. The working range of resistance thermometers is from -190 to $+500$°C. Some are designed to cover the range from -40 to $+100$°C. The resistance element usually has a value of the order of 25 ohms at 0.0°C, though some are made with a resistance of 2.5 ohms. The higher resistance units have a sensitivity of about 0.1 ohm/deg C, while those of lower resistance show a resistance change of about 0.01 ohm/deg C. Figure 80 is an illustration of a platinum resistance thermometer used in calorimetry work.

A special form of the Wheatstone network has been designed at the Bureau of Standards for precise temperature measurements. This assembly, which is commercially available, is known as the **Mueller bridge.** With this bridge temperature measurements can be made to 0.005°C.

The resistance type of thermometer is widely used in research and industrial work. For a detailed discussion of temperature measurements

by the resistance method the reader should consult "Methods of Measuring Electrical Resistance," Chap. XIII, by Dr. E. F. Northrup.

77. Measurement of Galvanometer Resistance. It is sometimes necessary to know the resistance of the winding of a current-indicating instrument such as a galvanometer. Lord Kelvin suggested an ingenious application of the Wheatstone network for the carrying out of such a determination. The circuit involved is sketched in Fig. 81. A comparison of this network with that of the original Wheatstone layout (Fig. 70) shows that the winding of the galvanometer replaces R_1. The

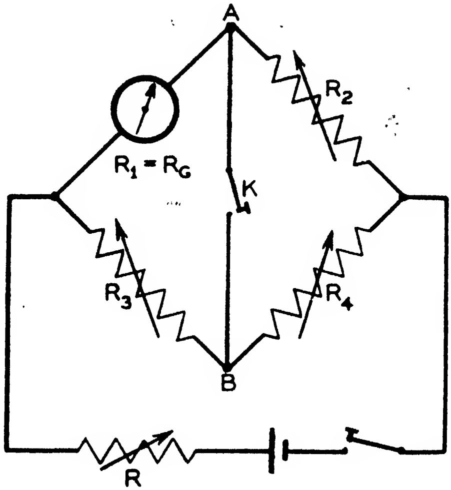


FIG. 81.—Measurement of galvanometer resistance.

branch AB contains only a short-circuiting switch K. By means of this arrangement the galvanometer, which is the resistance under test, also functions as the balance indicator. If and when the points A and B are at the same potential, the opening and closing of the key K **will not change the current flowing through the galvanometer arm.** In other words, the bridge will be electrically balanced and Eq. (103) will be applicable to the situation. The resistances R_2, R_3, and R_4 are adjusted, as usual, to secure a balance, *i.e.*, until no change in the galvanometer deflection occurs on opening and closing the key K. It will usually be necessary to insert a variable high resistance R, of the order of several thousand ohms, in series with the battery to prevent excessive deflections of the galvanometer.

78. Measurement of Electrolytic Resistance. One of the most important electrical measurements one is called upon to make is that of determining the specific resistance of an electrolyte. From elementary study

the reader is probably aware that an electrolyte is an aqueous solution of a salt, a base, or an acid. Usually the physicist or the physical chemist is interested in the **conductivity** of an electrolyte rather than its resistivity. But since conductivity is the reciprocal of resistivity, it is the latter which is actually measured. Later we shall consider the electrical behavior of electrolytes in some detail, but it will be convenient at this point in our

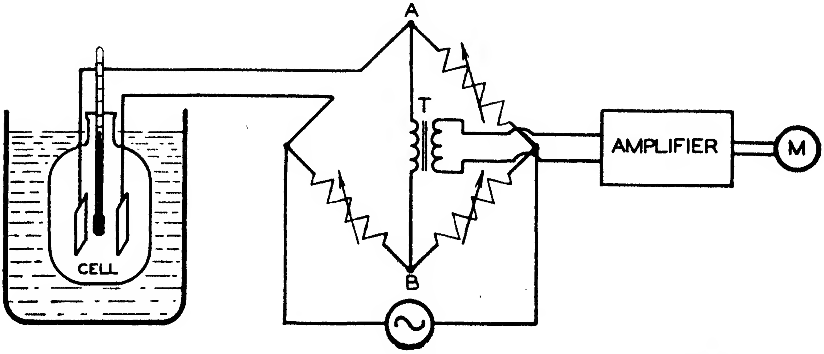


FIG. 82.—Schematic diagram showing improved method of measuring electrolytic resistance.

study to examine, briefly, the principles and procedure involved in the measurement of the resistance of conducting solutions.

The basic equipment involved in this type of resistance measurement consists of a well-made Wheatstone bridge, a special receptacle for holding the solution under test, a source of alternating potential, and a means of detecting an alternating current. The circuit setup is sketched in Fig. 82. The electrolytic cell, shown at the left of the diagram and in detail in Fig. 83, is usually of glass and designed to hold a few cubic centimeters of the electrolyte. Two platinum electrodes, each having an area of about 1 cm^2, are supported 1 or 2 cm apart and connected as one arm of the bridge. A constant-temperature bath surrounds the cell.

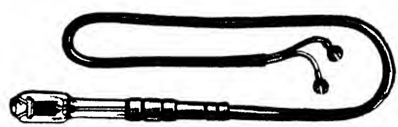


FIG. 83.—Electrolytic cell. (*Leeds & Northrup Co.*)

Direct current cannot be employed for this measurement, and for two reasons:

1. It would tend to decompose the electrolyte.
2. It would polarize the cell.

(These points will be elaborated upon in later chapters.) A source of alternating potential must therefore be available. Most up-to-date

laboratories are now equipped with a beat-frequency oscillator—an electronic device for producing alternating potentials at various frequencies, up to the order of something like 100,000 cycles per second (Sec. 238). Such an audio oscillator serves admirably as a source of a-c potential for electrolytic measurements and is greatly superior to the microphone hummer, formerly used. It is especially valuable since, by its use, measurements can be made over a wide range of frequencies.

In order to detect when the bridge was in balance, a set of headphone receivers or an a-c galvanometer was formerly used. An improved method consists in coupling the bridge circuit to an audio amplifier by means of a suitable audio transformer T. An "output meter" (a rectifying voltmeter) connected to the output of the amplifier will serve to indicate when the points A and B are at the same potential; at balance, its reading will be zero or some low minimum. Such an assembly is highly sensitive, serves to eliminate any personal equation, and makes possible the making of measurements in the presence of extraneous noise.

There are, of course, a number of chemical and electrical precautions which must be observed if accurate results are to be secured. These are outlined in various standard laboratory manuals.

Usually readings are made on a solution whose resistivity is known, and the conductivities of other solutions are then compared with that of the standard electrolyte. The relation involved would be $K_1/K_2 = R_2/R_1$; where K_1 and K_2 are the conductivities of the two electrolytes, and R_1 and R_2 the corresponding resistances as determined from the bridge measurements. A 5 per cent solution of NaCl (which is often used as a reference solution) would show, at room temperature, a specific resistance of the order of 14.9 ohms per centimeter cube. A 5 per cent solution of $CuSO_4$ would give a reading of about 52.6 ohms per centimeter cube. The accepted conductivity of a 5 per cent NaCl solution is 0.0742 mho-cm.

There are many other electrical measurements involving resistance determinations, but those outlined in the foregoing sections are representative of the more important cases.

79. The Potentiometer. Having outlined several standard methods of measuring resistance, we next turn our attention to the equally important subject of the measurement of emf and potential difference. In Chap. VI we described several instruments that may be used as potential-determining devices, but those instruments have certain inherent limitations as to range, sensitivity, and rapidity of operation. We are now to examine a principle, the application of which serves to provide a piece of equipment that finds wide and important uses in both research work and electrical engineering practice. Reference is made to the potentiometer method.

Poggendorff, in 1841, first described a "compensation method" for determining the emf of a primary cell without drawing appreciable current from it. Poggendorff's original procedure was later improved upon by Bois-Reymond (1862), Clark (1873), and others. In 1906, M. E. Leeds and E. F. Northrup described a greatly improved form of the assembly which Clark had designated as a **potentiometer.**

The fundamental principle involved in the operation of the potentiometer consists in **accurately balancing the emf under investigation against the drop in potential over a uniform resistance wire carrying a steady current.** The means by which this is accomplished may be understood by reference to the circuit diagram shown in Fig. 84. AB is a

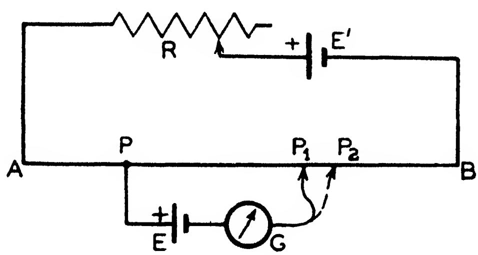


FIG. 84.—Illustrating fundamental principles of the potentiometer.

uniform resistance wire through which a current flows from the battery E'. G is a galvanometer, and E a standard[1] source of emf. By adjusting the current through A and B by means of the rheostat R, the drop (potential difference) between the contact points P and P_1 can be made equal to the emf of the cell E. When this condition obtains the galvanometer will read zero, since the emf and the drop are in opposition. It will thus be seen that the procedure constitutes a null method.

Having obtained a balance, the standard cell can be replaced by a cell whose emf it is desired to determine. With the current through AB accurately maintained at its original value a second balance is secured by adjustment of the movable contact on AB. In general the new balance will occur at some point, say, P_2. Since AB is uniform in resistance, and since the current through AB has not been changed since the first balance was secured, it may be said that

$$\frac{E_x}{E_s} = \frac{PP_2}{PP_1}. \tag{108}$$

Thus if E_s, the emf of the standard, is known, the emf E_x of the cell under

[1] A standard cell giving a fixed and accurately known emf can be assembled according to specifications laid down by the AIEE. A detailed description of such a cell is given in Sec. 99.

test can be computed from the reading of the lengths of the potentiometer wire utilized in the two cases.

It should be pointed out that the potentiometer may be utilized for the purpose of measuring **any** d-c potential difference; its use is not confined to the determination of the potential difference at the terminals of a source of emf. The fall in potential over a conductor through which a current is passing may become the unknown potential difference the magnitude of which is to be determined. The methods by which the potentiometer may be utilized in such cases are outlined in succeeding sections.

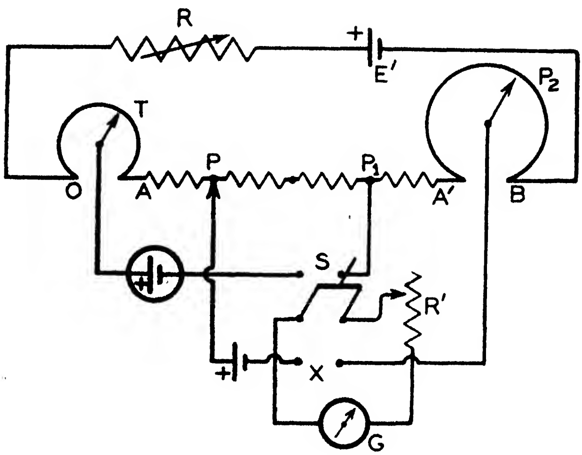


FIG. 85.—Schematic wiring diagram of a commonly-used type of potentiometer. (*Leeds & Northrup Co.*)

In practice, the potentiometer network is somewhat more involved than shown above. Figure 85 indicates the essential components constituting a modern form of the potentiometer. In this assembly the potentiometer resistor (corresponding to AB in Fig. 84) consists of **three** parts, *viz.*, OA, AA', and $A'B$. OA is a resistor variable in steps. Standard cells differ slightly in their emf value, being about 1.0183 volts. By adjustment of the contactor T on OA the drop over TP_1 can be made to correspond to that given by the particular standard cell being used. With this set to the proper value and the standard cell connected by means of the dpdt switch, the current in the entire resistor assembly OB is adjusted by means of R until balance is secured. In the Leeds and Northrup type K potentiometer, which is the unit most widely used in this country, a current of about 0.02 amp will give a suitable drop over the resistors provided. Once having established this current, by balancing against the standard cell, the cell under test is switched into the circuit and a

second balance secured by adjusting either one or both of the contactors P and P_2. AA' consists of a number of fixed resistors, and $A'B$ is an extended slide-wire form of resistor contact with which is made by means of a revolving drum. These resistance values are so chosen that with a current of, say, 0.02 amp flowing, the IR drop will be in volts, or fractions thereof. **Thus the apparatus reads directly in volts, and fractions of volts;** no computation is required, thus greatly facilitating the determination of the quantity sought.

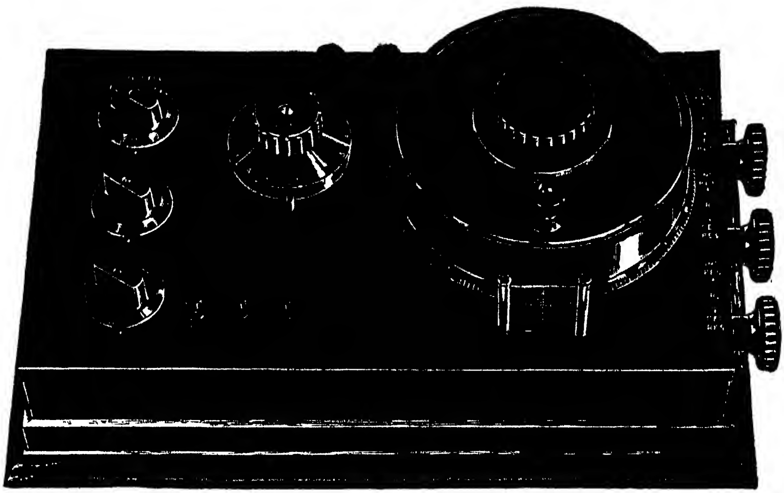


FIG. 86.—Commercial form of potentiometer, known as type K. (*Leeds & Northrup Co.*)

The emf developed by a standard cell is seriously changed if an appreciable current is drawn from the unit. Accordingly provisions are made in the potentiometer assembly whereby this contingency may be avoided. Referring to Fig. 85, it will be noted that there is a resistor R' in series with the galvanometer and the cell in use. This is commonly variable in two or three steps, having a maximum value of the order of 50,000 ohms. During the preliminary stages of the balancing adjustment all of this resistance is connected into the battery-galvanometer circuit, thus reducing the current drain on the cell to a negligible value. This protective resistance is cut out of circuit when the final balance is made.

With the type K potentiometer the basic range is from 0 to 1.61 volts. By making use of a shunting arrangement the range can be changed to 0 to 0.161. Within the higher range the potential may be read directly to 0.00005 volt and estimated to 0.00001 volt. When using the lower range, one can read directly to 0.000005 and estimate to 0.000001, or one

microvolt. In order to accomplish such results the resistance units entering into the construction of such a piece of equipment must be of the highest possible accuracy. Accordingly they are adjusted to an accuracy of ± 0.01 per cent. Figure 86 is an illustration of one model of the type K potentiometer.

If and when it becomes necessary to determine accurately potential differences greater than 1.61 volts, recourse is had to a so-called **volt box.** This is but another adaption of the potentiometer principle. A volt box consists of a high resistance divided into several steps, and so arranged that, say, a tenth, a hundredth, or a thousandth of the total resistance

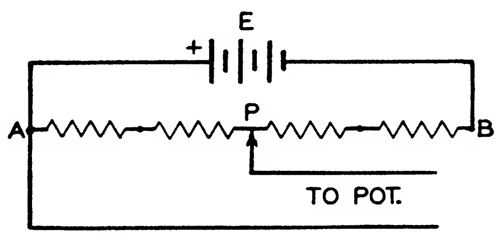


FIG. 87.—Circuit of multiplier or "volt box" used to extend the range of a potentiometer.

may be connected to the potentiometer. The source of emf or potential difference to be measured is connected to the terminals of the high resistance as shown in Fig. 87. By means of a contactor P a fraction of the total drop in potential between A and B is applied to the potentiometer and its value determined as outlined above. The reading of the potentiometer is then multiplied by 10, 100, or 1,000 as the case may be.

The potentiometer finds its widest usefulness in connection with the measurement of the emf developed by thermojunctions (Sec. 102), in the determination of pH values (Sec. 90), and in calibration of d-c meters. With the exception of the galvanometer, the potentiometer is probably the most widely used piece of electrical equipment. The student should therefore become **thoroughly** acquainted with the principles involved in its design, and with the technique connected with its applications. There are other types of potentiometers than the one described above, among which may be mentioned the Wolff, the Wenner, and the White; but the type K instrument is most commonly encountered.

80. Calibration of a Voltmeter. Reference was made above to the use of the potentiometer in the calibration of d-c meters. A d-c voltmeter may be calibrated by making use of the circuit setup sketched in Fig. 88. A high resistance AB (several thousand ohms) is connected across a source of emf whose total voltage is at least equal to the maximum range of the meter V being calibrated. This part of the network

functions as a volt box. A second high resistance $A'B'$ is connected across the meter terminals. A movable contact P makes it possible to impress any fraction of the emf of E on the meter, and the adjustment of contact P' provides a means whereby any small and convenient fraction of the potential difference at the terminals of the meter may be applied to the input terminals of the potentiometer. Thus, by applying the potentiometer method twice, a calibration of the meter may be made.

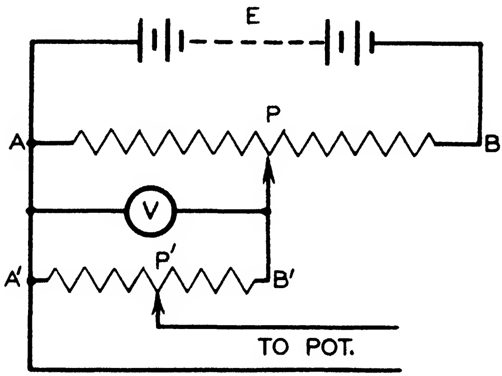


FIG. 88.—Circuit used in the calibration of a voltmeter by means of the potentiometer.

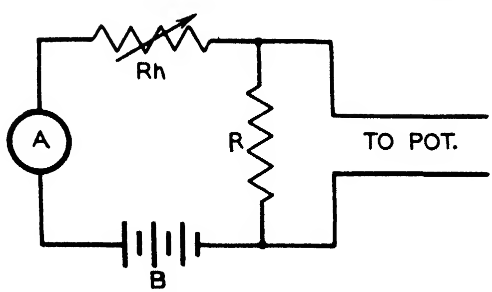


FIG. 89.—Calibration of an ammeter by means of the potentiometer.

81. Calibration of an Ammeter. Not only does the potentiometer constitute a means whereby potential differences may be conveniently and accurately determined, but it may also be utilized for the purpose of measuring current; thus making it possible to calibrate a d-c ammeter. Referring to the circuit diagram appearing as Fig. 89, B represents a storage battery or other constant source of emf. A is the ammeter to be calibrated, Rh a control rheostat, and R a resistance whose value is accurately known. This resistor is sometimes referred to as a **standard shunt.** By means of the rheostat, the magnitude of the current is given any desired value within the range of the meter and the corresponding potential drop over the standard resistor is determined by means of the

potentiometer in the usual manner. Knowing R, and having accurately measured V, the current I can be computed by means of Ohm's law. By taking a series of such readings the ammeter can be calibrated throughout its entire range. Standard resistors are available in values ranging from 0.1 to 0.001 ohm. A one-tenth-ohm shunt would be used when the maxi-

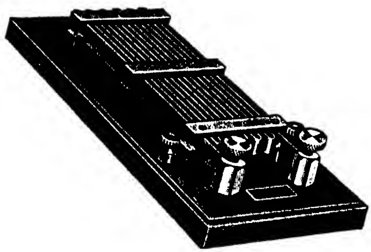


FIG. 90.—Typical standard shunt for use in the circuit shown in Fig. 89. *(Leeds & Northrup Co.)*

mum current is of the order of 15 amp; while a one-thousandth-ohm shunt would be designed to carry a current of several hundred amperes. Figure 90 shows a typical one-tenth-ohm standard resistor.

82. Measurement of Resistance by Means of the Potentiometer. From what has already been said about the potentiometer it is evident

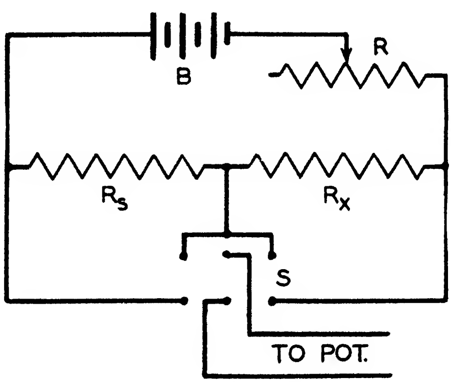


FIG. 91.—Circuit involved in the measurement of resistance by means of the potentiometer.

that this is a versatile piece of scientific equipment. Not only may the potentiometer method be applied to the measurement of potential differences and the calibration of d-c meters, but it may also be utilized for the purpose of measuring resistance. The procedure by which this end is accomplished will be apparent from an examination of the circuit sketched in Fig. 91. It will be seen that the resistor R_x under test

is connected in series with a standard resistor R_s and a source of emf B. The magnitude of the current through the resistor is controlled by the rheostat R. By means of the dpdt switch S alternate readings of the drop over R_s and R_x can be taken by means of the potentiometer. The same current flows through both resistors; hence the drop over the standard resistor will be R_sI and that over the unknown will be R_xI. Accordingly, we may write

$$\frac{V_x}{V_s} = \frac{R_xI}{R_sI},$$

which leads to

$$R_x = \frac{R_sV_x}{V_s}. \tag{109}$$

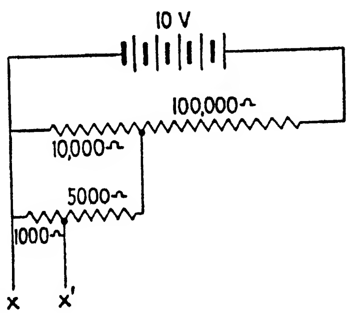
The accuracy of this method of measuring resistance will depend largely, on the accuracy of the calibration of the standard resistor.

PROBLEMS

1. If, in connection with some experimental procedure, you were in need of a potential difference of 0.001 volt, and there was available a dry cell (1.5 volts) and a 50,000-ohm resistor, explain how you would arrange conditions in order to obtain the desired potential difference. Would it be better to use a resistor of lower value?

2. Explain how you would arrange conditions so that a voltmeter which has a full-scale reading of 150 ma might be utilized as an ammeter having a full-scale value of 15 amp.

3. Assuming the conditions indicated in the accompanying sketch, what would be the potential difference between the terminals marked x and x'?



CHAPTER X

THERMAL EFFECTS OF THE ELECTRIC CURRENT

83. Joule's Law. There are three principal effects which result from the movement of electrons through a conductor; these effects are

1. The production of heat
2. The causing of chemical changes
3. The production of a magnetic field

The heating effect is always present whenever an electric current exists; the other two effects are produced if and when certain conditions obtain. Since the thermal effect is, in a sense, universal it will be discussed first.

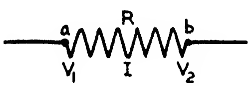


FIG. 92.—Work is involved in transferring electrons between two points in a conductor which are not at the same potential.

As previously pointed out, a part of the kinetic energy of the moving electrons is converted into heat, due probably to collisions between the electrons and the atomic nuclei. Whereas the process by which this energy transformation is brought about is not yet fully understood, we have, however, discovered the existence of certain important relations between the magnitude of the current involved and the heat developed. We shall now examine these relationships.

Consider a conductor ab (Fig. 92) the terminals of which are at the potentials V_1 and V_2. If Q statcoulombs of electricity (electrons) are transferred from a to b the work done will be given by the relation [Eq. (53)]

$$W = Q(V_2 - V_1).$$

Writing V for $V_2 - V_1$, we have

$$W = QV. \tag{i}$$

If Q is in statcoulombs and V in statvolts, W will be in ergs. If all of the energy expended in making the transfer is converted into heat it follows that

$$W = JH, \tag{ii}$$

where H is the heat in calories, and J the mechanical equivalent of heat, or 4.18×10^7 ergs/cal. By combining (i) and (ii) we may write the equivalents

$$W = JH = QV = ItV = I^2Rt.$$

From the above we have

$$H = \frac{I^2Rt}{J} \tag{110}$$

$$H = \frac{QV}{J} \tag{111}$$

$$H = \frac{IVt}{J} \tag{112}$$

$$W = IVt \tag{113}$$

$$W = I^2Rt. \tag{114}$$

In the above relations, I, R, and V are in esu, *i.e.*, in statamperes, statohms, and statvolts, respectively. The factor t is of course in seconds, H in calories, and W in ergs. Equations (110) to (114) are different forms of a relation known as Joule's law; they give the energy as a function of the several electrical factors and of time.

In order to use practical units in applying Eqs. (110) to (114) certain transformations must be made. Bearing in mind that the ampere is equivalent to 3×10^9 statamperes and that the ohm is equal to 9×10^{-11} statohms, Eq. (110) takes the form

$$H = \frac{I^2 \times 3^2 \times 10^{9\times 2} \times Rt}{4.18 \times 10^7 \times 9 \times 10^{11}} = 0.239I^2Rt, \tag{115}$$

which for most purposes may be written

$$H = 0.24I^2Rt;$$

where H is in calories, I in amperes, R in ohms, and t in seconds.

In a similar way, Eq. (111) may be changed thus

$$H = \frac{3 \times 10^9 QV}{300 \times 4.18 \times 10^7} = 0.239QV, \tag{116}$$

where H is in calories, Q in coulombs, and V in volts.

A transformation of Eq. (112) yields

$$H = \frac{I \times 3 \times 10^9 Vt}{300 \times 4.18 \times 10^7} = 0.239IVt, \tag{117}$$

where H is in calories, I in amperes, V in volts, and t in seconds.

When reduced to practical units, Eq. (113) gives

$$W = \frac{I \times 3 \times 10^9 \times Vt}{300} = IVt \times 10^7, \tag{118}$$

in which W is in ergs, I in amperes, V in volts, and t in seconds. The

above relation could take the form

$$W = \frac{IVt \times 10^7}{10^7} = IVt, \tag{119}$$

in which W is in joules.

Proceeding in a similar manner, we have for Eq. (114)

$$W = \frac{I^2 \times 3^2 \times 10^{9\times 2} \times Rt}{10^7 \times 9 \times 10^{11}} = I^2Rt, \tag{120}$$

where W is in joules, I in amperes, R in ohms, and t in seconds. Equations (115) to (120), inclusive, have many practical applications.

In the next chapter we shall examine various practical applications of the heating effect of the electric current. There are, however, many instances where this heating effect results in the wasteful dissipation of energy. One striking example of this is the I^2R loss, in power lines, dynamos, and motors. In electrical machinery not only is energy dissipated in the form of heat, but the thermal energy thus liberated may seriously damage the insulating material and, indeed, the conductors themselves. It is therefore important that a conductor of adequate cross section be selected to carry a given current under a given set of conditions. Open wiring can dissipate thermal energy more readily than when the conductor is enclosed or wound into a compact coil. The thermal conductivity of the insulating material is also a factor. The National Board of Fire Underwriters has established regulations governing the allowable current-carrying capacity of copper wire when used to connect a source of electrical energy to the electrical load.

Problem. An electric heater takes 5 amp at 110 volts. (*a*) How much heat is developed by the heating unit if operated for 2 hr? (*b*) How much energy is liberated as heat?

Solution. (*a*) Substituting in Eq. (117),

$$H = 0.24\,IVt = 0.24 \times 5 \times 110 \times 2 \times 3{,}600 = 9.54 \times 10^5 \text{ calories.}$$

(*b*) Making use of Eq. (119),

$$W = IVt = 5 \times 110 \times 2 \times 3{,}600 = 3.96 \times 10^6 \text{ joules.}$$

84. Electric Power. It will be recalled that the term power, in general, means **the time rate of doing work** or of expending energy. In the study of mechanics it is pointed out that, if energy is expended at the rate of 10^7 ergs—or 1 joule—per second, the power is said to be 1 watt. The watt is also utilized as a unit of power in connection with electrical computations. Let us proceed to derive a working relation which will enable one to compute the rate of energy expenditure (power) in an electrical circuit.

We may begin by examining Eq. (119) which is

$$W \text{ (joules)} = I \text{ (amperes) } V \text{ (volts) } t \text{ (seconds).}$$

This is equivalent to

$$\frac{W}{t} = IV$$

or

$$P = IV, \tag{121}$$

where P is the power in watts.

Another expression for power may be derived from Eq. (120), which is

$$W \text{ (joules)} = I^2 \text{ (amperes) } R \text{ (ohms) } t \text{ (seconds).}$$

Again dividing by t, we have

$$\frac{W}{t} = I^2R$$

or

$$P = I^2R, \tag{122}$$

where P is in watts.

Cases may arise in which it would be convenient to express power in terms of potential difference and resistance. Such a situation can be met by substituting the equivalent V/R for the term I in the last equation; thus securing the expression

$$P = \frac{V^2}{R}, \tag{123}$$

where P is in watts, V in volts, and R in ohms. The three relations for power which have been developed above **are applicable to any part of a circuit** as well as to the circuit as a whole. For instance, if one knows the current in a conductor that forms a component of a complete circuit and if we know, or can determine, the drop over that conductor, it is possible to compute the rate of energy consumption or dissipation in that particular conductor. This is an important point and should be clearly understood by the student.

It should also be noted that the relations embodied in Eqs. (121) to (123) hold only for those cases in which **the current is constant in both magnitude and direction.** In a-c practice another factor must be taken into consideration; that situation will be discussed in Chap. XX.

Electrical power may be expressed in other units than watts. Remembering that 746 watts equal 1 horsepower, we may write

$$1 \text{ electrical hp} = \frac{\text{electrical power in watts}}{746} \tag{124}$$

If energy is used for one second at the rate of one watt, the **amount of energy** involved will be one joule or one **watt-second.** Obviously the watt-second is a small unit, and accordingly a larger energy unit has been set up that is known as the kilowatt-hour. Thus we see that

$$1 \text{ kwhr} = 1000 \times 3{,}600 = 3{,}600{,}000 \text{ joules or watt-seconds.}$$

From fundamental relations it may be shown that

$$1 \text{ gm-cal} = 4.2 \text{ watt-sec}$$

It should again be stressed that the watt-second and the kilowatt-hour are **energy** units. Energy is used or dissipated at a certain rate (power) for a given length of time. The product of these two factors (rate and time) gives the total quantity of energy involved in any electrical operation. The kilowatt-hour is therefore employed as a basic unit in the merchandising of electrical energy.

Problem. A 6,600-volt, 20-mile, two-wire line supplies a 100-kw load. The copper conductors have a cross-sectional area of 26,250 cir mils. (*a*) What is the line loss in watts? (*b*) What must be the effective horsepower of the engine required to drive the generator to supply the load indicated?

Solution. (*a*) In order to compute the rate of energy dissipation in the line (line loss) it will be necessary to determine the line current, and also the line resistance. The line current can be found by applying Eq. (121)

$$I = \frac{P}{V} = \frac{100{,}000}{6{,}600} = 15.75 \text{ amp.}$$

The line resistance involves the use of Eq. (92).

$$R = \rho \frac{l}{a} = 10.4 \times \frac{20 \times 5{,}280}{26{,}250} = 40.9 \text{ ohms.}$$

The I^2R line loss can be computed by the aid of Eq. (122).

$$P = I^2R = (15.15)^2 \times 40.9 = 25{,}325 \text{ watts or } 25.33 \text{ kw.}$$

(*b*) Since the load is 100 kw and the line loss is 25.33 kw, the total power would be 125.33 kw. The effective horsepower which must be supplied by the prime mover would be given by Eq. (124).

$$\text{Hp} = \frac{125{,}325}{746} = 168.$$

PROBLEMS

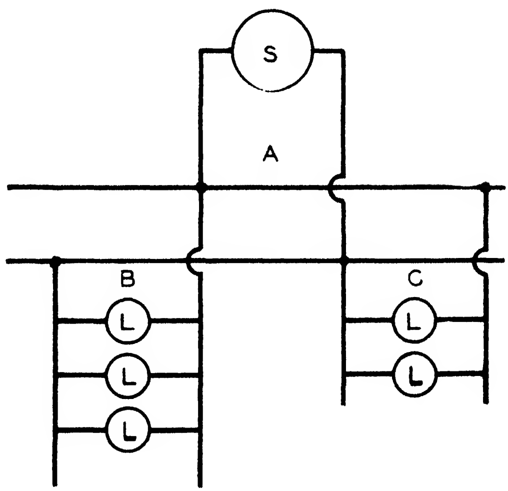
1. In a certain dehydrating process it is necessary to have available 10,000 Btu of heat during a period of 1 hr. If the heat is to be developed electrically from a 115-volt circuit what must be the resistance of the heating unit employed? What will be the magnitude of the current? If electrical energy can be had for 2 cents per kilowatt-hour, what will be the cost?

2. What quantity of electricity will be involved in the operation indicated in Prob. 1?

3. Soft-drawn copper has a resistivity of 1.724 microhm-cm at 20°C, while hard-drawn copper has a resistivity of 1.772 microhm-cm. The hard-drawn copper is commonly used for transmission lines because of its greater tensile strength (nearly 50 per cent greater). Compare the line loss in two cases for lines that have conductors of equal diameter and length.

4. A residence is supplied by a 100-ft two-wire copper line using No. 10 conductors. Without changing the service leads, an electric stove is installed. If the maximum current taken by the stove is 25 amp, to what extent will the service voltage be reduced?

5. The accompanying sketch represents the wiring plan of a house. Branch *A* supplies a stove using 1,000 watts, branch *B* is a lighting circuit carrying ten 50-watt lamps, and branch *C* supplies two 100-watt lamps. The voltage at the



house fuse box is 115. Compute: (*a*) The total current; (*b*) the total wattage; (*c*) the resistance of each branch, neglecting the line resistance; (*d*) the total resistance; (*e*) the cost of operating the total load for 1 hr if the energy rate is 4 cents per kilowatt-hour.

6. Compare the cost of heating 25 gal of water from 60 to 110°F when heated by gas and by electricity. Assume that the gas costs \$1 per 1,000 ft^3 and yields 1,000 Btu/ft^3, and that the electrical energy costs 4 cents per kilowatt-hour.

7. A two-wire 10,000-volt power line, 50 miles in length, delivers 1,000 kw to an electrical load. If the wire is No. 0 (diameter = 325 mils) what will be the line loss in kilowatts? Suppose the voltage is raised to 100,000 volts; what would be the loss? Compare the monthly monetary loss in the two cases, on the basis that it cost 0.5 cent per kilowatt-hour to produce the electrical energy.

8. A d-c electric motor is rated at 100 hp when operated at 220 volts. If such a motor is supplied by a two-wire 100-ft pair of feeders, size 600,000 cir mils,

what potential difference must be maintained at the input end of the mains? Would there be any advantage in using a 550-volt motor in such a case?

9. An electric motor draws 10 amp at 115 volts, and operates at an efficiency of 95 per cent. What is the output of the motor in horsepower? What is the heat loss in watts? How many Btu are developed per hour? What energy in joules is consumed if the motor is operated for 8 hr?

10. What should be the resistance of a heating coil which will raise the temperature of 1 gal of water from 20 to 100°C in 30 min? The heater is to be operated on a 115-volt circuit.

CHAPTER XI

APPLICATIONS OF THE THERMAL EFFECT

85. Electric Furnaces. Furnaces which utilize the electric current as a heating agency are in extensive research and commercial use. One reason for this is that they are, in general, more efficient than those of the fuel-fired type; and further, the temperature is under complete control. An additional feature is the freedom, when desired, from contaminating factors such as carbon, etc. Electric furnaces assume many forms, depending upon the particular use for which they are designed.

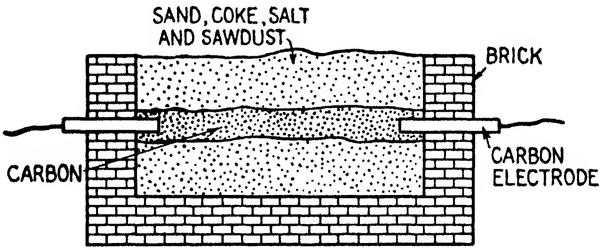


FIG. 93.—Diagrammatic sketch of the resistance furnace in which carborundum is made.

It would be beyond the scope of this volume to present a review of all types; however, a description of two or three typical units will be given.

Electric furnaces may, for convenience, be classified into three general types: **the resistance furnace, the arc furnace**, and **the induction furnace.**

One of the first resistance furnaces to come into extensive use was that developed by Acheson in 1891 for producing the abrasive material known as carborundum. A cross-sectional view of the carborundum furnace is shown in Fig. 93.

A certain carborundum furnace consuming 2,000 kw has a length of 40 ft and is 5 ft across. The outer walls of the furnace are made of loosely assembled bricks and are temporary in nature—the side walls being taken down after each "firing" in order to remove the charge. The end walls are permanent, however, and carry the large water-cooled carbon electrodes, two electrodes (3 ft by 10 in.) being connected in parallel at each end.

In charging the furnace, the mixture, consisting of sand, coke, sawdust, and salt, is filled in until the furnace is about half full. A trench-like path 2 ft wide and 1½ ft deep is then made lengthwise in the mixture

between the electrodes and into this is placed the "core." This core, consisting of carefully prepared coke, serves as a conductor and has a resistance of the order of 0.03 ohm.

When the core material has been laid in place, the remainder of the mixture is filled in and rounded over on top, the top not being covered with brick. Owing to the relatively high resistance of the core when cold it is necessary to apply a voltage of 300 at the beginning. The operating potential difference, after stable thermal conditions obtain, is about 200 volts, the voltage being regulated in order to maintain constant energy input. A stable working condition is established in from one to two hours.

Alternating current is used, the energy being supplied by means of step-down transformers. The high temperature developed (about 2000°C) results in a chemical reaction between the sand and carbon as shown by the equation

$$SiO_2 + 3C \rightarrow 2CO + SiC,$$

SiC being the carborundum. The function of the sawdust in the mixture is to make the mass more porous and thus facilitate the escape of the carbon monoxide gas. As this gas escapes from the sides of the furnace it is set on fire and burns during the time the furnace is in operation. It is interesting to note in passing that for every 1,000 lb of carborundum produced, 1,400 lb of gas are given off. The salt reacts with any iron present to form the chloride.

The heating continues for 36 to 40 hr, after which the furnace is allowed to cool. It is then opened by taking down a part of the brick walls and the charge is removed. In a furnace of the size described, from 5 to 8 tons of carborundum are secured from each "heat." In one installation located near Niagara Falls 27 such furnaces are in operation.

Incidentally it may be mentioned that graphite is produced in a furnace of the above-described type. In fact, if the temperature in a carborundum furnace be allowed to exceed 2240°C decomposition of the silicon carbide begins, silicon being driven off as a vapor and graphitic carbon remaining. In a furnace designed for the production of graphite, anthracite coal and sand constitute the mixture for the charge. Graphite is widely used as a constituent of lubricants. It is also employed in the manufacture of electrodes for certain electrolytic processes, its electrical conductivity being about four times as great as that of the amorphous variety of carbon.

Comparatively small electric furnaces of the resistance type, designed for laboratory use, are constructed by winding a heat-resisting tube, or vessel, with resistance wire such as nichrome or molybdenum, and inclos-

ing the unit in a jacket of thermal-insulating material. Furnaces of this type are used, for example, in connection with the calibration of pyrometers and platinum thermometers.

As an example of the arc type of furnace we may take a unit which is used in the production of high-grade steel. Figure 94 is a diagrammatic sketch of the Héroult steel furnace, the drawing being more or less self-explanatory. This unit is generally designed for use with a three-phased current (Sec. 140), and hence three electrodes are used, each of which is connected to the power supply. Furnaces of this type are designed to

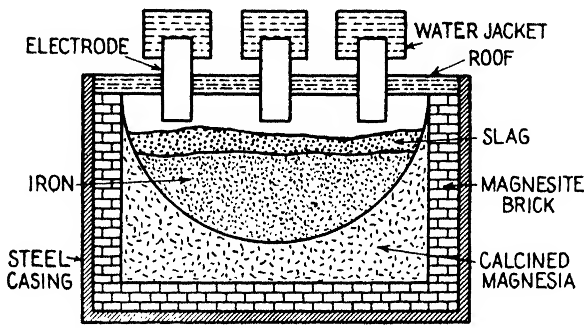


FIG. 94.—Showing plan of Héroult arc furnace used in the production of steel.

handle from 15 to 20 tons of steel at one charge. A 15-ton unit absorbs 2,250 kw at 110 volts between phases. Such furnaces are frequently charged with molten metal from a Bessemer converter or open-hearth furnace. A mixture of lime and iron oxide is introduced on top of the molten metal and an arc is formed between the electrode and the slag, and thence to the steel beneath, the current leaving the furnace by arcing from the steel to the slag and in turn to the other electrode. The slag serves as an oxidizing agent for the purpose of removing phosphorus. When tests show that phosphorus has been sufficiently removed, the furnace, which is so arranged mechanically that it may be rotated, is tilted and the slag removed. Other slag consisting of lime, fluor spar, and coke dust is then added. The heat developed by the electric arc transforms a part of this slag into calcium carbide, which in turn is decomposed, liberating a certain amount of carbon which is absorbed by the steel. From time to time the steel is tested in order to determine when the desired grade has been secured. By means of this furnace low-grade Bessemer steel is converted into a high-grade product relatively free from oxygen, sulphur, and phosphorus, and containing a definite and predetermined amount of carbon. A furnace of the type just described will yield a steel ingot of about 8 tons per charge. From 1½ to 2 hours are required for a "heat."

In 1916 Dr. Edwin F. Northrup invented a special form of induction furnace in which a high frequency alternating current is utilized to develop the thermal effects. It will be shown in a subsequent discussion (Sec. 131) that electric currents are developed in any conducting material that is subjected to a magnetic field that varies in intensity. Such "induced" currents may, under suitable conditions, assume large values. This is particularly true if the magnetic field varies (alternates) at a high frequency (several thousand cycles per second).

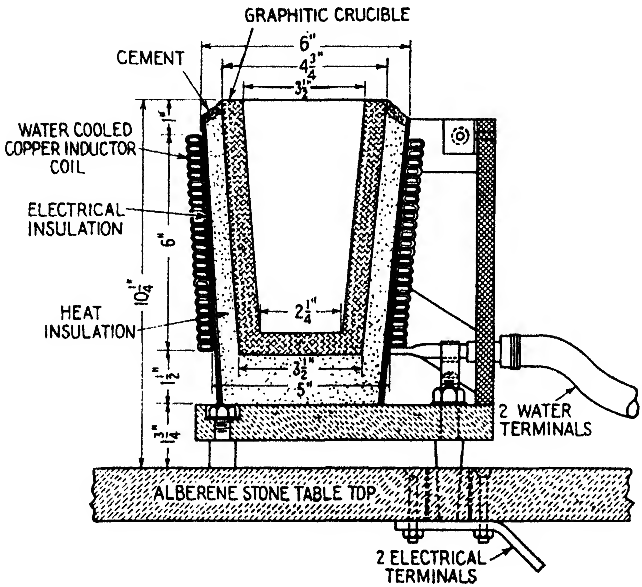


FIG. 95.—High frequency induction furnace. (*Ajax-Electrothermic Corp.*)

In the h-f induction furnace designed by Dr. Northrup the receptacle containing the material to be melted is surrounded by a so-called inductor coil (Fig. 95) from which it is electrically and thermally insulated. This inductor winding consists of a flattened copper tube through which water flows for the purpose of cooling the conductor. When a h-f alternating current from a suitable source (Sec. 238) is passed through the inductor large induced currents will be developed in any conducting material in the crucible, and these induced currents will cause rapid and efficient heating of the material. If a nonconducting substance, such as glass, is to be heated the crucible is made of carbon and the induced currents are set up in the receptacle itself. Temperatures as high as 2800°C may be produced, and the temperature is under complete control at all times.

Another interesting and valuable feature of the Northrup furnace is

that the induced currents which cause the heating also cause a movement or circulation of the molten material, which, in many cases, is an important part of the melting process.

Aside from temperature control, perhaps the most important advantage of this type of furnace is that the melting of any material may be carried out without chemical contamination of the substance being worked. In fact, it is entirely practicable to melt a substance *in vacuo*, or in the presence of some particular gas.

Furnaces of this type are made in sizes ranging in capacity from a few ounces to several tons. The frequency of the alternating current used depends upon the nature of the work for which a particular unit is designed and ranges from 500 to 50,000 cycles/sec. The power consumed in the Northrup furnace ranges from 3 kw in the smallest laboratory unit to 1,500 kw in the larger commercial installations.

In recent years the use of induction heating has greatly increased, particularly since high-power electron tubes have become more generally available. As we shall see later, such tubes may be made to replace dynamos as generators of alternating currents of practically any frequency. Using tube generators it is possible to heat-treat various metals, either superficially or throughout the body of material, and in many cases to do it much more expeditiously than by any other available method. The piece of metal to be heated is placed within an helical winding carrying an alternating current having a frequency of the order of 100 kilocycles to 20 megacycles. The eddy currents (Sec. 131) developed in the metal quickly heat the body of material to any desired temperature. The principle involved is exactly the same as that employed in the Northrup furnace. Surface hardening, annealing, and various other metallurgic processes may be rapidly carried out, and at a relatively low cost. Electronic h-f heating units as large as 200 kw are now commercially available. This method of utilizing the heating effect of the electric current is being rapidly extended.

86. Electric Welding. The thermal effect of the electric current finds wide and increasing application in electric welding. Referring again to the relation expressing Joule's law [Eq. (115)] we note that the heating effect of the current varies as the **square** of the current and as the **first power** of the resistance in the circuit. If, then, arrangements are made to handle very heavy currents intense heating effects may be produced either throughout the entire body of the metal, or locally on the surface. Further, the heat so produced, except for oxidation, is "clean," which is an important consideration in the process of welding. There are two principal methods of electric welding: the Thompson process, and arc welding.

In the **Thompson process,** the two pieces of metal (as for example, two streetcar rails) are placed end to end and, by means of heavy leads clamped to them near their ends, a heavy current from a step-down transformer (Sec. 147) is sent across the junction. Because the resistance at the point of contact of the rails is relatively high heating occurs at that point. When the ends of the two members have been brought to the proper temperature, pressure is applied in order to force the ends into intimate contact and the current is turned off. Alternating current of 40 to 60 cycles is used, and the step-down transformer employed in the process has only one or two turns of heavy copper in the secondary. The current in the welding circuit is of the order of 25,000 amp, depending upon the nature and size of the weld.

In **arc welding** the heat is developed locally, usually at or near the surface of the metal being worked. For example, in "bonding" streetcar rails that are not welded, heavy flexible copper conductors are welded to each rail near the end. This is accomplished by connecting the rail to the positive terminal of a d-c supply and the negative lead to a portable carbon electrode. To effect a weld the carbon electrode is brought into contact with the metals to be welded and the resulting arc heats the materials at the point where the weld is desired. Currents of the order of 200 amp are employed, at a voltage of about 50. Specially designed d-c dynamos are employed for this type of work. Very extensive use is made of this method of welding, and for some classes of work it is replacing acetylene gas welding. The control of the current in this type of welding is now effected by means of electronic tubes.

87. Electric Lighting. In 1800, Sir Humphry Davy, the English scientist, discovered that a brilliant electric discharge took place when the electric current from a number of primary cells was caused to pass between two carbon terminals that were first in contact and then slightly separated. When the carbon electrodes were in a horizontal position the luminous vaporized carbon between the terminals assumed a bow-like form, due to the heated air, and hence the term "arc" was used by Davy to designate the phenomenon. After Faraday's later discoveries which served as a basis for the development of the dynamo, Davy's discovery led to the development of the modern electric arc. Notwithstanding the many improvements incorporated in the modern arc lamp, this unit is now more or less obsolete as a source of illumination, except for certain special purposes. The spectrum of the carbon arc is rich in violet and ultraviolet light and is accordingly sometimes used as a source of this type of radiation. The temperature of the carbon arc is exceedingly high, being about 3500°C—literally the "hottest place on earth." If the arc is enclosed in an insulating housing an even higher temperature can be

secured. The positive terminal of a d-c arc is the point of highest temperature. All elements, including tungsten, have melting points below the temperature of the arc. For this reason the arc may be utilized as a means of melting any of the metals.

There is one further characteristic of the arc which it may be well to note in passing. Reference is made to the fact that it exhibits what is referred to as a **negative resistance.** This characteristic is also indicated by saying that it has a **falling characteristic.** The meaning of these two expressions is that Ohm's law does not hold for the current through the arc; as the current increases **the potential drop across the arc becomes less.** Any conducting agent which behaves in this manner also possesses certain other important and useful characteristics which we shall have occasion to examine later.

A lamp which, in certain respects, might be classified as an arc lamp is the **mercury-vapor** unit, commonly known as the **Cooper-Hewitt lamp,** after its originator. This lamp consists of a glass tube from 1 to 4 ft in length and about 1 in. in diameter. Mercury forms one electrode while a piece of metal or graphite serves as the other (Fig. 96). The arc is "struck" by tilting the tube (normally supported in an oblique position) until the mercury forms a thin thread connecting the two electrodes. When the tube is returned to its original position an arc is formed, thus vaporizing some of the mercury. The arc then continues to operate between the electrodes through the metallic vapor as a path. Both d-c and a-c models are in use. Means are provided for automatic starting. The lamp operates at an efficiency of about 12 lumens/watt and has a life of some 4,000 hr.

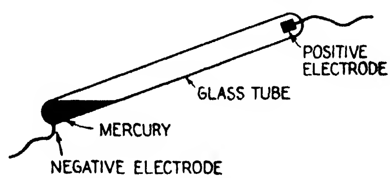


FIG. 96.—Cooper-Hewitt mercury vapor lamp.

This lamp yields a light in which greenish blue predominates, only a faint red line being present. Consequently, the Cooper-Hewitt lamp cannot be utilized where exact color values are important. The lamp has found extensive use in warehouses, factories, and other similar places but is now being replaced by the fluorescent unit. The light from the mercury lamp is rich in actinic rays and has therefore been used to some extent as an artificial light in photographic work.

The efficiency of the mercury arc may be increased by operating the unit at a higher temperature and pressure. In practice this is made possible by making the enclosure of quartz instead of glass, a short unit being the result, and one giving a light more nearly white. Quartz, however, transmits ultraviolet light, and hence, when utilized for ordi-

nary purposes, it becomes necessary to enclose the unit in glass, thus filtering out the short wave lengths. The enclosed quartz mercury-vapor lamp has an efficiency of 26 lumens/watt.

Ultraviolet light possesses bactericidal properties. The fact that the unscreened quartz lamp may serve as a source of intense ultraviolet light is taken advantage of for the sterilization of water, for therapeutic purposes, and for the sterilization of the air in schoolrooms and hospitals. When used as a sterilization agent the water to be treated is caused to flow slowly over one or more such lamps and is thereby sterilized. In the case of air the mercury lamp is located in an air duct through which air is caused to enter a room.

Small models of both the glass and quartz mercury lamps are made for use in scientific laboratories as a source of short-wave radiation. Two pools of mercury in an exhausted tube serve as the electrodes; and the unit operates on a d-c potential of about 25 volts. By means of suitable optical filters the blue and other radiation from a mercury arc can be eliminated leaving a single green spectral line whose wave length is 5,461 Å. Such a source of monochromatic light finds wide use in connection with spectroscopic measurements.

Another source of illumination involving the movement of charges through attenuated gases is commonly referred to as the **neon lamp.** This is a modern version of the original Moore tube lamp. This lamp consists of a glass tube several feet in length within which is neon, or some other inert gas, at a low pressure. When an alternating potential difference of the order of 2,000 volts is applied to the lamp terminals the attenuated gas is rendered luminous. The electrical process by which electrical energy is converted into radiant energy under these circumstances will be discussed in a later chapter. Whereas in the case of the mercury lamp the radiation consists mainly of the shorter wave lengths, the radiation from a neon lamp consists largely of the lower frequencies—those which we designate as red and orange. Such luminous tubes are widely used for advertising and display purposes. The efficiency is of the order of 17 lumens/watt.

In recent years another form of vapor lamp known as the **sodium-vapor lamp** has been developed. This unit consists of a glass enclosure containing a small amount of inert gas and a supply of metallic sodium. When subjected to a potential difference of the order of 20 volts at its terminals, a discharge will be initiated through the inert gas, much as in the neon lamp. The heat developed by this discharge will in time vaporize some of the sodium, which in turn will become the medium by means of which the arc is maintained. The efficiency of the sodium-vapor lamp is remarkably high, being about 65 lumens/watt. Its radiation is

nearly monochromatic. It exhibits the characteristic yellow of the sodium spectrum. However, the visual effect, particularly for outdoor lighting, is pleasant; and this type of lighting unit is coming into use as a means of highway illumination. Sodium arcs are made for operation on both alternating and direct current.

The most widely used illuminant is, of course, the **incandescent lamp.** Originally developed about 1880 by Swan in England, and Edison in this country, the incandescent unit has undergone many improvements. The original carbon filaments gave an efficiency of about 1.3 lumens/watt and had a relatively poor spectral output. Still it was a marked improvement over the oil lamp as a source of visible radiant energy. In 1911, Dr. Coolidge developed a process whereby tungsten could be produced in ductile form, and in 1913 Dr. Langmuir found that the introduction of an inert gas into a lamp bulb made it possible to operate a filament at a high temperature without excessive evaporation of filament material. These two discoveries resulted in the development of the modern incandescent lamp. A metallic filament operated at a relatively high temperature (about 3000°K) made it possible to take advantage of Wein's displacement law and thus secure a spectral radiation giving a relatively high electrooptical efficiency. Modern incandescent lamps show an efficiency ranging from approximately 11 lumens/watt for the 50-watt size, to at least 20 lumens/watt for the 1,000-watt units.

The work of Coolidge, Langmuir, and others in connection with the development of the modern incandescent lamp constitutes one of the outstanding achievements in applied science. However, if one takes into account both the physiological and electrical factors involved, the best incandescent lamp shows a **visual** efficiency of less than 2 per cent. Recently a marked improvement in over-all efficiency has been brought about by the introduction of the **fluorescent lamp.**

Reference was made above to the spectral limitation of the mercury-vapor lamp. It has of course been long known that certain chemical substances exhibited fluorescence when irradiated by ultraviolet light. Such substances act as optical transformers; they receive short (invisible) wave length radiation and transform it into visible radiation having a relatively broad spectral distribution. The modern low-voltage fluorescent lamp consists of a glass tube 14 to 47 in. in length, depending on the energy rating. The inside of this tube is coated with an adherent fluorescent material called a phosphor. One such substance consists of zinc and beryllium silicate. An electrode, in the form of a short coated filament, is sealed into the tube at each end. A small amount of argon gas (to facilitate starting) and a drop of mercury are enclosed in the tube. The operation of the lamp will be understood by reference to Fig. 97.

Upon closing the switch S the filaments are heated to incandescence, thus liberating electrons from their chemically coated surfaces (Sec. 188). These electrons, moving under the action of the alternating field, cause ionization (Sec. 204) of the argon gas, thus establishing a current through the tube. A few seconds after the switch is manually or automatically opened a potential surge of sufficient magnitude to start the discharge through the tube between the operating electrodes will be set up by the reactor. Once in operation the ballast reactor (choke) serves to limit the magnitude of the current through the tube. The ultraviolet

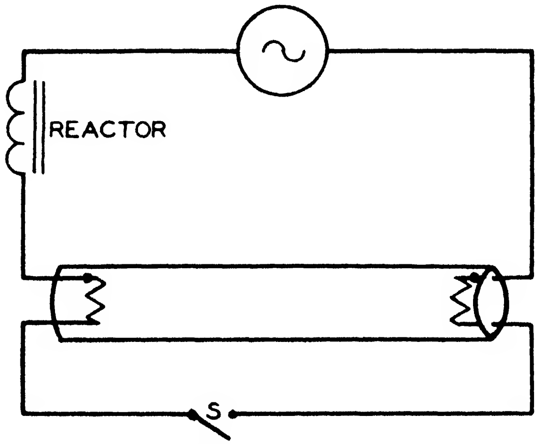


FIG. 97.—Diagram of the circuit involved in the operation of the fluorescent lamp—switch-starting type.

radiation (chiefly the wave length 2,537 Å) from the ionized mercury causes the phosphor to emit visible radiation showing a continuous spectrum. Low-voltage fluorescent lamps draw about 0.25 amp; they operate at a temperature of 40°C, and thus approach the ideal of a "cold light." By proper selection of phosphors a number of colored lights can be produced, including a close approach to "daylight" radiation. The starting and control components consume 2 or 3 watts, but the over-all efficiency is materially higher than in the case of incandescent lamps. The 15-watt daylight unit, for instance, operates at an efficiency of 30 lumens/watt. In order to secure the same quantity of light flux having the same spectral characteristics a 150-watt incandescent unit would be required. Thus the power ratio is 10:1 in favor of the fluorescent lamp. The most spectacular gain in efficiency is, however, in the production of colored light for use in decorative and architectural lighting. In this field the advance in efficiency is indeed remarkable. For instance, a tube yielding green light shows an efficiency of 60 or more lumens/watt,

as against 0.3 lumen/watt for corresponding filament lamps. This represents an efficiency gain of approximately two hundredfold. In the case of a blue fluorescent unit the gain over incandescent bulbs is about fifty-fold. Low-voltage fluorescent lamps will not operate efficiently in surroundings where the ambient temperature is in the region of zero °C. The operating life of fluorescent lamps is comparable with the filament-type light source. Figure 98 shows a circuit arrangement by which electrode heating and potential augmentation are simultaneously

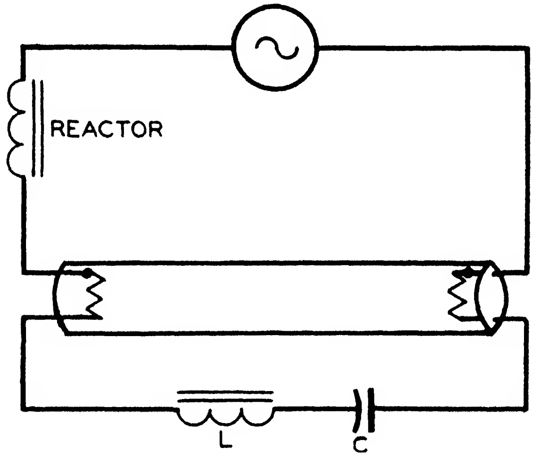


FIG. 98.—Circuit diagram of the automatic-starting type of fluorescent lamp.

attained. This network also makes the starting of the lamp automatic once the line switch is closed. The reactor L and the condenser C, in parallel with the tube, form a semiresonant circuit and thus serve to accomplish the above-indicated ends. Just how this comes about will be discussed later.

Fluorescent lamps are also now made for operation at voltages ranging from 1,000 to 15,000 supplied from the secondary of a transformer. As in the low-voltage tubes, mercury vapor is most commonly employed as the exciting agent, though neon gas is used to some extent, particularly where ambient temperatures are low. The high-voltage tubes are smaller in diameter than the low-voltage units, ranging in size from 8 to 35 mm. More than 100 ft of tubing may be connected in series and operated from a single transformer. The efficiency depends upon the color of the light being produced, the dimensions of the tube, and the operating conditions. The range is from 18 to 60 lumens/watt, the highest efficiency being attained in the case of the green. Cold white shows an efficiency of 26 lumens/watt. These tubes draw from 8 to 250 ma of current, the most

commonly used size taking 60 ma. Tube life is said to range from 3,000 to 10,000 hr. High-voltage fluorescent tubes are used chiefly for outdoor illumination, such as advertising and architectural displays. For further details regarding this advance in illuminating engineering the reader should consult a comprehensive review of the subject entitled "Fluorescent Lighting Advances" in the June, 1941, issue of *Electrical Engineering*.

CHAPTER XII

CHEMICAL EFFECTS OF THE ELECTRIC CURRENT

88. Electrolysis. In dealing with the transfer of electrons along a metallic conductor it was pointed out that **only** electrons take part in the migration—the atoms, as entities, do not undergo physical displacement, and there is no chemical change. There are, however, substances in which the mechanism of electrical conduction is quite different. Reference is made to aqueous solutions of salts, bases, and acids. Such solutions (called electrolytes) may act as conductors, but the process by which electrical conduction occurs involves a physical movement of charged atoms and groups of atoms. Any discussion of the process of conduction in solutions involves a consideration of the nature of solutions.

It has been established by the work of Arrhenius and other investigators that when an acid, base, or salt is dissolved in water a considerable number of the molecules at once undergo dissociation into positively and negatively charged entities which are called **ions.** This situation might be represented by the expression

$$MR = M^+ + R^-,$$

where, in general, M represents an atom of a metal or hydrogen, and R a group of atoms acting as a transient chemical entity. M^+ and R^- represent positive and negative ions, respectively. In solution these ions probably move about in an irregular manner. Possibly some of the water molecules are also dissociated thus,

$$H_2O \rightleftarrows H^+ + OH^-.$$

The modern view in regard to electrolytes is to the effect that in the case of most salts, strong acids, and strong bases, the percentage dissociation is materially greater than originally thought by Arrhenius. It is now generally held that, even in the solid state, a salt consists of positive and negative ions bound together by electrostatic forces. In other words, **ionization** probably obtains even in solid salts, particularly if the salt is in the crystalline form. The reader is cautioned against confusing ionization and dissociation; the terms are not synonymous. When a salt is dissolved in water, the ions constituting a given molecule are still attracted toward one another by the electrostatic force due to their opposite charges. This interionic electrostatic force is, however, very much less in a solution than in a solid, the chief reason being that the dielectric

constant of water is high, and, therefore, by Eq. (2), the force of attraction will be comparatively small. In concentrated solutions there is probably a definite pairing of ions, or possibly a frequent interchange of ionic partners. It is thus seen that although ionization may be more or less complete, **dissociation will not be complete** except in the case of very dilute solutions. In dilute solutions the ions are so far apart that the **interionic electrostatic force is exceedingly small**—and this is the significance of the term **dissociation.** Under such circumstances both the positive and negative ions will be attracted by any suitably charged bodies which may be introduced into the electrolyte.

To be more specific, if n molecules of sodium chloride are dissolved in m molecules of water, the situation could be represented thus:

$$n\mathrm{NaCl} + m\mathrm{H_2O} \rightleftarrows m\mathrm{H_2O} + x\mathrm{Na^+} + x\mathrm{Cl^-} + (n - x)\mathrm{NaCl}.$$

In the case of hydrochloric acid, the case would be

$$n\mathrm{HCl} + m\mathrm{H_2O} \rightleftarrows m\mathrm{H_2O} + x\mathrm{H^+} + x\mathrm{Cl^-} + (n - x)\mathrm{HCl}.$$

In these two cases sodium and hydrogen atoms, with their single positive charge, constitute the positive ions; while the chlorine atoms, plus the negative charge on each, constitute the negative ion. The elements above cited are monovalent and hence carry a single charge, *i.e.*, an excess or a shortage of one electron.

If we were to make a solution of a salt like copper sulphate, for instance, we would find that the situation differed somewhat from the examples cited above. The dissociation reaction would be

$$n\mathrm{CuSO_4} + m\mathrm{H_2O} \rightleftarrows m\mathrm{H_2O} + x\mathrm{Cu^{++}} + x\mathrm{SO_4^{--}} + (n - x)\mathrm{CuSO_4}.$$

Copper is divalent and hence, in the ionic state, carries two elemental charges. Likewise the radical SO_4 acts as a divalent group and accordingly has two excess electrons. Trivalent atoms carry three unit charges, etc. In other words, the ions carry charges proportional to the valence of the atom, or group of atoms, that acts as a chemical unit. The reader is cautioned against confusing the terms **ion** and **electron.** The former refers to an atom or group of atoms together with its associated charge (an excess or shortage of electrons); the latter refers to the natural unit quantity of electricity.

If some copper sulphate ($CuSO_4$) is dissolved in water and the resulting solution is connected to a source of electrical energy, as shown in Fig. 99, dissociation will take place, as pointed out above. Assuming, for the moment, that the electrodes which dip into the solution are of some neutral material, such as platinum or carbon, copper will be deposited on the negative terminal and the radical SO_4 will appear at the

positive electrode, **the movement of the ions through the electrolyte being due to the electrostatic attraction of the charged electrodes.** This process is known as **electrolysis.** Each copper ion will acquire two electrons from the negative electrode (called the cathode) and thus becomes an electrically neutral atom of copper, and will in general attach itself to the electrode. Each negative ion (SO_4^{--}) will deliver two electrons to the positive electrode (the anode) and become an uncharged chemical

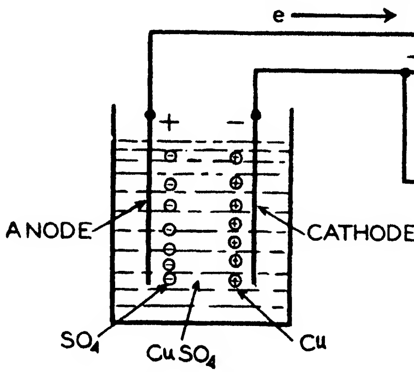


FIG. 99.—Electrolytic process.

entity having a very strong affinity for hydrogen. It will accordingly unite at once with two atoms of hydrogen from the dissociated water to form sulphuric acid, thus

$$2H + SO_4 \rightleftarrows H_2SO_4.$$

If this process continues the entire copper content of the electrolyte will be deposited on the cathode, and the electrolyte will consist of sulphuric acid only. If the anode were, say, copper, instead of some nonactive element, the SO_4 group would combine with the copper to form copper sulphate and thus tend to maintain the ionic concentration of the electrolyte. This is the basic process utilized in the electrolytic refining of the metals.

In the electrolytic process those ions that, under the influence of an electrostatic force, move toward the anode are called **anions**; and those which move toward the cathode are designated as **cations.** During the process of the transfer outlined above, the ions function as the carriers of the electric charges. The total current is composed of the charges carried by **both** types of ions **within the electrolyte,** yet the net result is an interchange of electrons. **Free electrons do not exist in the electrolyte.** In the **external circuit** we have a **movement of electrons only,** and this electronic current is maintained by the potential difference at the terminals of the external source of emf.

In general, nonaqueous solutions are nonconducting. Fused salts, however, do conduct the current.

The conductivity of electrolytes is a function of the ionic concentration, the degree of dissociation, the ionic velocity, the temperature, and other factors. A complete discussion of this aspect of the subject is beyond the scope of this volume. It will suffice to say that electrolytes, like metallic conductors, follow Ohm's law, and that their temperature coefficients of resistance are negative.

89. Faraday's Laws. The researches[1] carried out by Faraday in 1833–1834 led him to formulate the following very useful generalizations or laws, the first of which may be stated thus: **The weight of the material liberated at either of the electrodes is proportional to the total quantity of electricity which passes through the electrolyte.**

It will be recalled [Eq. (89)] that

$$Q = It.$$

If then, as stated above, $M \propto Q$, we may write

$$M = zQ = zIt, \tag{125}$$

where M is the weight of an ion set free at the electrode, I the current, t the time, and z a proportionality constant that is known as the **electrochemical equivalent.** By making Q unity in the last equation above, we may define the electrochemical equivalent as **the weight of an element in grams which is deposited by the transfer of unit quantity of electricity,** *i.e.*, by 1 coulomb.

Faraday's second law is to the effect that **when a given quantity of electricity is passed through two or more electrolytic cells in series, the quantity of ions liberated at a given electrode will be proportional to the chemical equivalents of the particular elements involved.**

It will be recalled that

$$\text{Chemical equivalent} = \frac{\text{atomic weight}}{\text{valence}}.$$

For example, if conditions are as illustrated in Fig. 100, and we pass 1 coulomb of electricity through the two cells in series the relative amounts of copper and chromium deposited will be in proportion to the ratio between the chemical equivalent of the two elements, thus

[1] The extremely primitive experimental facilities available at the time of Faraday's epoch-making investigation make that research one of the most remarkable in all scientific history. The student is urged to read Faraday's original account of his historic research. It is to Faraday that we owe the terminology now used to describe electrolytic processes.

$$\frac{\text{Weight of Cu deposited}}{\text{Weight of Cr deposited}} = \frac{63.57/2}{52.01/3} = \frac{31.785}{17.50}.$$

It is thus evident that, for a given quantity of electricity, more copper than chromium will be plated out of solution.

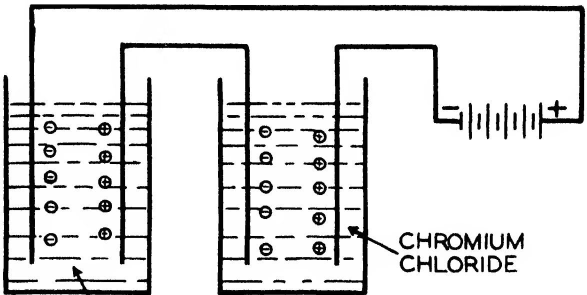


FIG. 100.—Illustrating Faraday's second law.

Following is a table giving the electrochemical constants of a number of elements encountered in the more common electrolytic processes.

Element	Atomic weight	Valence	Chemical equivalent	Electrochemical equivalent (z) (gm per coulomb)
Aluminum	27.0	3	8.96	0.00009316
Chlorine	35.46	1	35.46	0.0003674
Copper	63.57	1	63.57	0.0006588
Copper	63.57	2	31.78	0.0003294
Chromium	52.01	3	17.50	0.0001796
Gold	197.2	3	65.21	0.0006812
Hydrogen	1.008	1	1.008	0.0000104446
Iron	55.84	2	27.92	0.0002894
Iron	55.84	3	18.61	0.0001929
Lead	207.20	2	103.60	0.0010736
Nickel	58.68	2	29.34	0.0003041
Oxygen	16.0	2	8.0	0.00082902
Platinum	195.23	2	97.6	0.0010115
Silver	107.88	1	107.88	0.0011179
Sodium	22.997	1	22.997	0.0002383
Tin	118.7	2	59.35	0.0006150
Tin	118.7	4	29.7	0.0003075
Zinc	65.37	2	32.68	0.0003387

In dealing with electrolytic processes one encounters a term frequently used in chemical calculations. We refer to **gram equivalent. The gram equivalent of any element is the amount of the element, in grams, which is equal to the atomic weight divided by the valence.** Bearing in mind our previous definition of chemical equivalent it is evident that the gram equivalent of an element is **numerically** equal to its chemical equivalent. For instance, 29.34 gm of nickel constitutes the gram equivalent of that particular element.

It is important to calculate the quantity of electricity required to deposit electrolytically, or plate out, the gram equivalent of an element. From the foregoing definitions it may be stated that the quantity of electricity q required to deposit a gram equivalent of an element will be given by the expression

$$q = \frac{\text{gram equivalent}}{\text{electrochemical equivalent}}. \tag{126}$$

Taking silver as an example we have

$$q = \frac{107.88}{0.0011179} = 96{,}494 \text{ coulombs.}$$

This means that 96,494 coulombs of electricity will deposit 107.88 gm of silver from an electrolyte. If we were to go through the same calculations for any other elements it is obvious that we would obtain the same numerical ratio for q. It may therefore be said that the passage of 96,494 coulombs of electricity through an electrolyte will result in the deposition of a gram equivalent of any element. The number 96,494 is one of the important constants of nature; it is called the **faraday.**

Knowing the numerical value of the faraday, we can calculate the magnitude of the charge carried by each atom in the electrolytic process. From the atomic theory it is known that 1 gm equivalent of silver contains 6.032×10^{23} atoms of the element. This is Avogadro's number N. It follows, therefore, that in the process of depositing 1 gm equivalent of silver from one of its salts each silver ion carries an amount of electricity given by the relation

$$\begin{aligned} e &= \frac{96{,}494}{6.023 \times 10^{23}} = 1.601 \times 10^{-19} \text{ coulomb/atom} \\ &= 4.803 \times 10^{-10} \text{ statcoulombs/atom.} \end{aligned}$$

This figure, it will be recalled, **is the same as that for the elemental charge of electricity**—the electron. It is therefore evident that each silver atom, in the example cited, shows a positive charge equivalent to one

electron. The complementary negative ion, NO_3^- for instance, would carry one electron. Any other univalent element would give us the same result.

If one were to carry through the calculation in the case of a divalent element, such as nickel, it would be found that the charge being transported would be $2(4.803 \times 10^{-10})$ statcoulombs; and in the case of a trivalent element, such as chromium, the ionic charge would be

$$3(4.803 \times 10^{-10}) \text{ statcoulombs.}$$

The corresponding negative ions would carry two and three electrons, respectively.

In concluding our discussion of Faraday's laws, it may be pointed out that a knowledge of the first law makes possible the establishment of an electrochemical definition of unit current. Transposing Eq. (125) we get $I = M/zt$. Taking a particular metal, and making t 1 sec, it may be said, for instance, that **an ampere is that constant current which will deposit silver from an aqueous solution of silver nitrate at the rate of 0.0011179 grams per second.** This is, in fact, a definition of the **international ampere.** The electrochemical conditions under which the standard test is to be made were carefully prescribed by the London conference of 1908. For less accurate results copper may be used instead of silver. Later, we shall examine another, but equivalent, definition of the ampere.

90. Applications of the Electrolytic Process. An accurate and rapid quantitative determination of the chemical constituents of a compound may be carried out by means of the electrolytic process. Suppose that it is desired to determine the percentage of copper in a certain sample of brass (Cu + Zn). A representative sample of the metal to be tested would be dissolved, say in sulphuric acid. This compound salt solution would be placed in a platinum dish which had previously been carefully weighed. A platinum spiral or gauze immersed in the electrolyte would serve as an anode (positive terminal) and the platinum dish would function as the cathode (negative electrode). If a potential difference of about 0.34 volts is applied, deposition of the copper will take place. Exner has shown that the deposition process may be greatly expedited by agitating the electrolytes while electrolysis is taking place. Rotating the anode or boiling the solution will accomplish this end. Complete deposition on the platinum receptacle will take place in a few minutes, as indicated by chemical tests. When this has occurred the cathode is washed, dried, and again weighed. The difference in the two weighings will give the weight of the copper present in the sample, and the percentage composition can be readily computed. By using a voltage of 0.76 instead of 0.34 the zinc content can be plated out in a similar manner.

The example cited above is representative of the procedure widely employed in electrochemical analysis.

The extent to which the electric current is used in electrochemical processes in the manufacturing industries is very great and is constantly increasing. These applications include the electroplating of copper, silver, nickel, chromium, and gold; the refining of various metals; and the manufacture of a number of widely used chemical compounds. The plants which utilize electricity in this manner are located at or near points where cheap hydroelectric power is available. In this country many of them are situated near Niagara Falls where several hundred thousand kilowatts are utilized for such purposes. We have already discussed the fundamental principles on which these various commercial applications rest. It only remains to describe briefly a few of the typical commercial electrochemical processes.

In the plating of copper, silver, gold, nickel, and chromium the electrolyte usually consists of a complex salt. For example, in nickel plating the double nickel-ammonium sulphate [$NiSO_4 \cdot (NH_4)_2SO_4 \cdot 6H_2O$] is used.

Copper plating is a somewhat simpler process than is the case in the deposition of gold and silver. An acid solution of copper sulphate may be utilized as an electrolyte. In the printing industry copper electrotype replicas are produced on an extensive scale. In making a copper electrotype of a printed page or photograph a wax or gelatine mold of the original is first prepared. This impression is then dusted over with finely powdered plumbago. Powdered iron is then sprinkled over the carbon dust and the surface thus prepared is immersed in a solution of copper sulphate. As a result of the chemical reaction copper is deposited on the graphite. After washing, the prepared wax plate is placed in a solution of copper sulphate and made the cathode; a piece of pure copper serves as an anode. A current density of from 0.9 to 1.3 amp/ft^2 is used. When copper has been deposited to the desired thickness the metallic shell is separated from the wax form and "backed" with type metal or other material. The production of master phonograph records involves essentially the same process.

Refining of copper has become an extensive industry. It consists of the electrolytic deposition of copper from an electrolyte of acidulated copper sulphate. Crude copper bars or plates are suspended in large vats and connected as anodes. Sheets of pure copper serve as cathodes at the beginning of the process. By the proper regulation of the current and voltage copper having a purity of 99.95 per cent is deposited on the cathodes. Copper of this purity is needed for use in the manufacture of wire which is to be utilized in the electrical industry. In order to carry on the electrolytic refining of copper on a large scale, large d-c generators

are required, the vats being arranged in a series-parallel circuit A total current of several thousand amperes may be employed in a given refining plant.

One of the most extensive commercial applications of the electrolytic process is that connected with the production of the metal aluminum. In 1886 Hall in the United States and Héroult in France developed an economical method of utilizing electric current for the separation of this important metal. Aluminum cannot be electrolytically deposited from an aqueous solution of its salts. However, Hall and Héroult found that it is possible to deposit the metal from a molten mixture of cryolite and

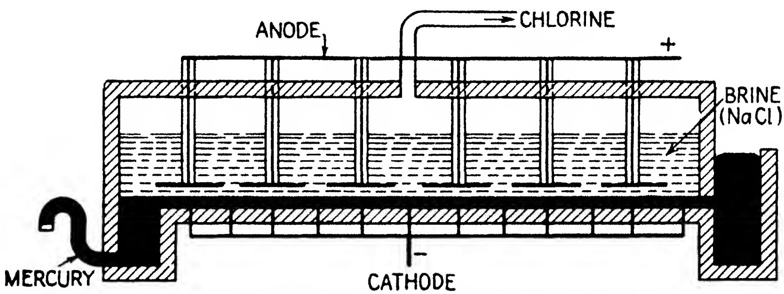


FIG. 101. -Kellner-Solvay cell as used for the making of sodium hydroxide by the electrolytic process.

alumina (Al_2O_3). The mixture of bauxite and cryolite is put into carbon-lined iron pots, the carbon serving as a cathode. An emf of 5.5 volts is applied and a current of 10,000 amp is passed through the mixture in each pot. The electric current serves the double purpose of maintaining the cryolite in a melted condition and of separating the aluminum oxide into aluminum and oxygen. The metal appears at the cathode in a molten state, collecting at the bottom of the receptacle. The oxygen which is liberated at the anode combines with the anode, which is carbon, thus gradually consuming that electrode. The hot aluminum is drawn off from the pots and cast into ingots, after which it is worked into wires, sheets, and rods.

As an illustration of the manufacture of chemical compounds by the electrolytic process, mention may be made of the production of caustic soda. Various electrolytic cells or organizations have been devised whereby it is possible to produce economically this chemical compound by electrical means. Several of the commercially successful cells involve essentially the same basic principle, viz., the use of mercury as the cathode. One form of such a cell is illustrated diagrammatically in Fig. 101, this particular organization being known as the Kellner-Solvay cell.

In this unit a thin layer of mercury is caused to flow across the floor of a large cement trough. Above the mercury a stream of brine (a solution of NaCl) is caused to move continually. A number of platinum wire-mesh anodes are supported just above the layer of mercury that serves as cathode. The application of an emf of 5 volts results in the separation of the sodium at the cathode which reacts to form an amalgam with the mercury. Provision is made whereby the amalgam is separated by gravity from the general mercury stream and subsequently brought into contact with water, thereby forming sodium hydroxide (NaOH), the product sought. The chlorine which appears at the anode escapes into the chamber above the brine and may be drawn off and utilized as a by-product. The energy efficiency of the electrochemical process just outlined is about 45 per cent, 1 kg of NaOH being produced for each kilowatt-hour of electrical energy consumed.

Electrochemical processes are being rapidly extended in a number of industrial fields.

PROBLEMS

1. In calibrating an ammeter by the electrochemical method, a deposit of 0.5 gm of silver was deposited on the cathode. If the current passed for 15 sec, what was the value of the current?

2. If electrical energy costs 1 cent per kilowatt-hour, what will it cost to plate out 100 lb of copper from an electrolyte in which the valence of copper is 2? Assume that an emf of 2 volts is used.

3. A current of 2 amp flows through two electrolytic cells in series, one of which contains copper sulphate and the other nickel-ammonium sulphate. How long must the current flow in order that 25 gm of copper shall be deposited? How much nickel would be plated out during the same period?

4. From the electrochemical constants, compute the atomic weight of oxygen.

5. On the assumption that the density of hydrogen is 0.00009 gm/cc, how many cubic centimeters of that gas would be liberated by the passage of 2 amp through a dilute sulphuric acid solution during a period of 10 min?

6. Suppose that a current of 1 amp passes from an iron pipe to the surrounding moist earth for a period of 30 days. How much iron will have been electrolytically removed from the pipe?

7. A faraday of electricity will liberate 1.008 gm of hydrogen. On that basis how many grams of chlorine will it liberate? How many grams of gold will it deposit?

8. Assuming that the area of the chromium trimmings on an automobile totals 5 ft^2, and that the plating is 0.1 mm thick, what will be the cost of the electrical energy involved in the operation? The density of chromium is 6.9 gm/cm^3, and the electricity costs 2.5 cents per kilowatt-hour. The plating is done at 2 volts.

CHAPTER XIII

PRIMARY AND SECONDARY CELLS

91. Electrode Potential. Intimately associated with the questions involved in electrolytic processes is the subject of the theory of electrode potentials. When immersed in an electrolyte why do certain elements show a potential which is different from that of the electrolyte? As a result of the careful investigations carried on by Arrhenius, Hittorf, Ostwald, Nernst, and others on the properties of solutions, particularly electrolytes, the explanation of this important phenomenon which has come to be accepted more or less generally may be summarized as follows.

Let it be assumed that a piece of chemically pure zinc, for instance, is immersed in a solution of one of its salts, say zinc sulphate ($ZnSO_4$). As previously pointed out (Sec. 88), some of the sulphate molecules will have undergone dissociation, thus giving rise to positive zinc ions and negative sulphate ions. We have seen that these ions appear to move about in an irregular manner. Some of the sulphions (SO_4^{--}) sooner or later will reach the immediate vicinity of the zinc electrode. The metal constituting the electrode has a tendency to unite with sulphions to form zinc sulphate, **the zinc going into solution in the ionic state.** The attraction of a negative ion for a surface zinc ion is greater than the sum of the interatomic attractions. (This tendency is referred to by the term **solution pressure.**)[1]

Since the zinc goes into solution as a positive ion, the electrode is left with an excess of two electrons and hence manifests a negative charge. This process may be represented by the expression $Zn \rightarrow Zn^{++} + 2e$; the two electrons remain on the electrode. The life of any newly formed zinc sulphate molecule is probably only transient; dissociation presumably takes place at once. Any zinc ions thus formed will be attracted to the now negatively charged electrode. There is thus established what may be referred to as an **electrical double layer.** The presence of this positively charged region of the electrolyte tends to prevent additional zinc ions from passing into solution. Thus we have two oppositely directed forces acting within the solution; one (solution pressure) tending to

[1] Nernst appears to have been the first investigator to use this term. In his writings it refers to a property by virtue of which all metals and hydrogen tend to pass into solution in the ionic state. Later studies have led to the conclusion that it is not a true pressure, as the term is commonly employed. It is possible that this expression covers several forces that, for convenience, may be expressed as a pressure.

cause ions to go into solution, and a second (electrostatic repulsion) acting to inhibit the first mentioned process.

There is also a third force involved in the situation, viz., a definite osmotic pressure, or the tendency of the zinc ions to go out of solution. Electrochemical **equilibrium will come to obtain when the sum of the electrical stress and the osmotic pressure equals the solution pressure.** This condition is arrived at so quickly, and involves so little material, that no chemical change can be detected when a metal is immersed in one of its salts.

The electrochemical process just outlined has, however, resulted in establishing a difference of potential between the metallic electrode and the electrolyte at the surface of contact. In other words, our electrode has acquired a definite potential with respect to the electrolyte, and this potential is spoken of as the **electrode potential.** In the case of zinc in a solution of one of its own salts the electrolyte in the immediate vicinity of the electrode is at a potential of 0.485 volt **above** that of the electrode itself.

In the case cited above (zinc in a zinc salt solution) the solution pressure of the metal is **greater** than the osmotic pressure of the ion. If, however, we take such a combination as copper in a copper sulphate solution we have a situation in which the solution pressure is **less** than the osmotic pressure. Under these circumstances there is a tendency for the metallic ion to **go out of solution,** and thus be deposited on the electrode. The electrode will accordingly acquire a positive charge. This interchange of charges may be represented by the expression $Cu^{++} + 2e \rightarrow Cu$, the **two electrons being supplied by the electrode.** The positively charged electrode will repel other positive ions, thus tending to inhibit the deposition of additional charges. As soon as the electrical stress comes to equal the difference between the solution pressure and the osmotic pressure, equilibrium will obtain, and further electronic transfer will cease. As a result of the loss of one or more positive ions to the electrode, the electrolyte in the immediate vicinity of the electrode will have an excess of electrons; in other words, it will be at a **lower** potential than the electrode, the potential difference being 0.621 volt. In this, as in the previous case, the size of the electrode has no bearing on the magnitude of the electrode potential.

The question at once arises as to whether the nature of the electrolyte (a salt of the electrode element) has any bearing on the magnitude of the electrode potential. Newmann has very thoroughly investigated this question. He found that the sulphates and chlorides, when used as electrolytes with any given metal, yielded slightly different values of electrode potential, the sulphate giving the higher value. Some metals, nevertheless, have the same electrode potential regardless of the salt

solutions in which they are immersed. If, however, the electrolyte does not consist of an aqueous solution of a salt of the electrode element, the electrode potential may depend upon the nature of the electrolyte. In fact, the potential of a given electrode with respect to one electrolyte may be positive, while the same electrode may show a negative potential when immersed in some other solution.

In addition it may be added that the degree of dissociation of the salt forming the electrolyte affects the magnitude of the electrode potential. In general the potential shows an increase with an increase in dissociation.

It has also been found that electrode potential is a function of the ionic concentration. This is to be expected because of the fact that, as we have seen above, osmotic pressure is one of the factors having a bearing on the electroequilibrium conditions at any given electrode; and osmotic pressure is known to be directly proportional to the concentration of the dissolved salt. Indeed, in any given case, it is possible to compute electrode potential from a knowledge of the osmotic pressure. In fact, the relation between electrolytic solution pressure, osmotic pressure, and electrode potential can be given mathematical form.

Following a line of argument suggested by Nernst, let us derive such a relation.

In dealing with the questions having to do with electrode potential we may think of the surface of an electrode as functioning as a semipermeable membrane **for ions**; the electrode and the electrolyte acting as the two mediums. For the purposes of this discussion, electrolytic solution pressure may also be thought of as being similar in character to osmotic pressure. Van't Hoff has shown that one can compute the magnitude of osmotic pressure, at least in dilute solutions, by assuming that the solute acts as a perfect gas having the same volume as that occupied by the solution. On such an assumption the ions would follow the general gas law, viz.,

$$PV = RT, \tag{i}$$

where P is the osmotic pressure, V the volume involved, R the gas constant, and T the absolute temperature.

The work done in causing a gram equivalent of the electrode material to pass into solution in the ionic state would be given by the relation

$$W = \int_{P_2}^{P_1} P\,dV, \tag{ii}$$

where P_1 is the solution pressure and P_2 the osmotic pressure. But from (i)

$$V = \frac{RT}{P}.$$

Substituting this value for V in (ii) we get

$$W = RT \int_{P_2}^{P_1} \frac{dP}{P} = RT \log_e \frac{P_1}{P_2}. \tag{iii}$$

But the energy expended in doing this work has been transformed into electrical energy. The total energy involved in this process would be given by EnF, where E is the potential difference established between the solution and the electrode, n the valence of the ion involved, and F the faraday. We may therefore equate this expression to the energy relation given by (iii) and get

$$EnF = RT \log_e \frac{P_1}{P_2},$$

which reduces to the form

$$E = \frac{RT}{nF} \log_e \frac{P_1}{P_2}. \tag{127}$$

This is a form of the well-known Nernst equation; it gives the electrode potential in terms of certain definite, known constants, the electrolytic solution pressure, and the corresponding osmotic pressure. Later we shall apply this general relation to the case of several important special electrodes.

We may conclude our present discussion with the observation that **the seat of electrode potential is at the surface of separation of the electrode and the electrolyte** and that its magnitude depends upon the relative values of the solution pressure of the element and the osmotic pressure of the ion involved.

92. Reference Electrodes. The measurement of absolute electrode potential presents a number of difficulties. Because of this it has been agreed to adopt a particular electrode as an arbitrary standard or base from which to compute all other electrode potentials. By international agreement a **hydrogen electrode,** in a normal solution of hydrogen ions, has been taken to be such a reference base; and its emf is **considered to be zero** at all temperatures. This combination is frequently referred to as the **normal hydrogen electrode.**

With regard to electrochemical reactions hydrogen behaves as a metal and will function as a reversible electrode. In order to make an electrode of hydrogen it becomes necessary to provide a means whereby the gas may be kept in contact with the electrolyte. This is accomplished by utilizing a small piece of platinum, whose surface is coated with platinum black (finely divided platinum). A platinum surface thus prepared will occlude a large quantity of hydrogen, and can be made to function as a **gas electrode.** When in contact with an electrolyte containing hydrogen ions some of the occluded hydrogen changes from the gaseous to the

ionic state. In short, if constant gas pressure is maintained, such an electrode functions in a manner analogous to an electrode consisting of a piece of metal dipping into a solution of one of its salts. The platinum metal as such is not a factor in the case. If such an electrode is saturated with hydrogen at atmospheric pressure, and partly immersed in a solution containing hydrogen ions (say, $H^+ + Cl^-$) of such a concentration that it has 1 gm molecule of hydrogen ions per liter, a definite and reproducible potential will be developed—which potential **is arbitrarily taken to be zero,** under the conditions mentioned. **Actually,** the potential difference between a hydrogen electrode and a solution containing 1 gm molecule of hydrogen ions per liter is known to be -0.277 volt, and not zero. The hydrogen electrode as introduced by Bottger,[1] and later improved and simplified by Hildebrand,[2] consists of a glass housing which serves to protect the platinum and confine the atmosphere of hydrogen. The electrode itself (Fig. 102) may consist of a piece of platinum foil P of about 1 cm² area or a short piece of platinum wire, the connection to which is sealed into an inner glass stem as shown. This component is in turn inclosed in an outer glass tube provided with a side tube through which gas is introduced, and apertures near the lower end through which the gas may escape.

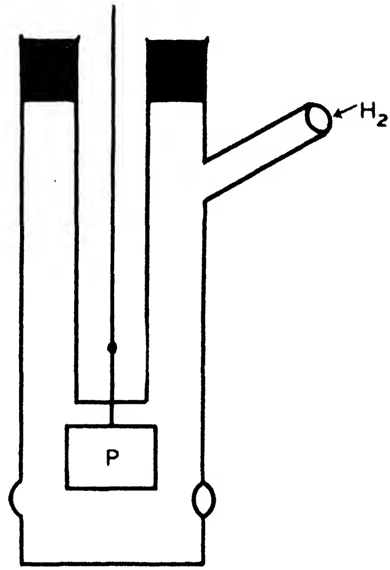


Fig. 102.—Hydrogen electrode.

If, then, we desire to determine the potential of a given electrode, an assembly such as that sketched in Fig. 103 may be utilized. The metal M under investigation is partially immersed in a solution containing one of its salts. By means of a tube B containing a salt solution (and known as a salt bridge) the electrolyte S is electrically connected to a hydrogen electrode as shown. The difference of potential between the connections a and b is then measured by means of a potentiometer. Since the potential of the normal hydrogen electrode is taken to be zero the potentiometer reading gives the potential difference between the solution S and the electrode M **in terms of the normal hydrogen electrode.**

Employing the comparison method outlined above the electrode

[1] *Z. physik. Chem.*, **24**, 253 (1897).

[2] *J. Am. Chem. Soc.*, **35**, 849 (1913).

potentials of the various elements, in a normal salt solution, have been carefully determined. Certain representative values are given in the following list, arranged in their order in the electrochemical series. The data listed in the table have been compiled from various recent sources. The marked values are somewhat doubtful.

Electrode Potentials in Volts

Element	Ion	Potential referred to hydrogen electrode (25°C)
Potassium	K	+2.924
Magnesium	Mg	+2.40*
Aluminum	Al	+1.70*
Manganese	Mn	+1.10*
Zinc	Zn	+0.762
Iron	Fe	+0.441
Cadmium	Cd	+0.401*
Nickel	Ni	+0 231
Lead	Pb	+0.122
Iron	Fe	+0 045*
Hydrogen	H	0.000
Antimony	Sb	−0.10*
Copper	Cu	−0.344
Mercury	Hg	−0.799
Chlorine	Cl	−1.358
Gold	Au	−1.50*

Since the absolute electrode potential of hydrogen is known to be −0.277 volt, as indicated above, it is possible to compute the absolute potential of any electrode by making the necessary correction. For instance, in the case of zinc, the absolute potential would be

$$(0.762) + (-0.277),$$

or 0.485 volt. For copper we would have $(-0.344) + (-0.277)$, or −0.621 volt.

While the hydrogen electrode still remains the ultimate standard of reference, for reasons of experimental convenience this form of electrode has been largely replaced in practice by the **calomel electrode.** Mercury in contact with mercurous chloride (Hg_2Cl_2) constitutes this electrode combination. Such an electrode is constructed as shown diagrammatically in Fig. 104. The potassium chloride (KCl) in the connecting tube forms

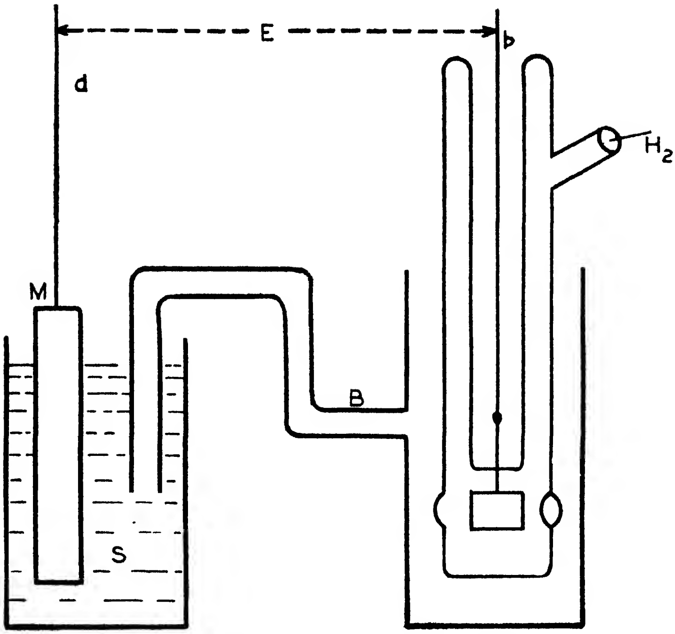


FIG. 103.—Determination of electrode potential by means of a hydrogen electrode and the potentiometer.

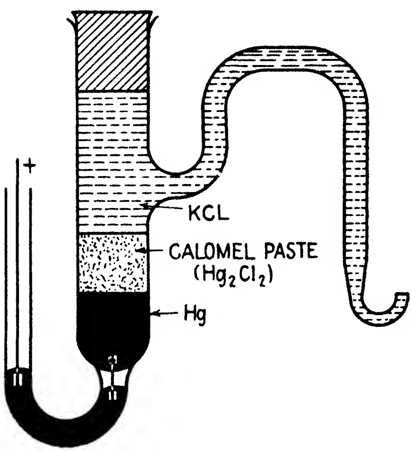


FIG. 104.—Older form of calomel cell.

FIG. 105.—Modern form of calomel cell (*Leeds & Northrup Co.*)

a "salt bridge" which serves to minimize any possible potential difference at the junction of the two electrolytes. The mercury electrode is given various forms, depending upon the use to which it is to be put. One type currently in use consists of a small and very simple form, as sketched in Fig. 105. In this model the salt bridge consists of the capillary space between the lower and upper sections of the outer glass housing.

The potential of the calomel electrode varies with the KCl concentration and with the temperature of the cell. Below is given the potential, referred to the hydrogen electrode, for the three common types of such electrodes at the temperatures ordinarily encountered. These data were supplied by the Leeds and Northrup Co. Because it is more easily pre-

Calomel electrode	Potential, volts		
	At 20°C	At 25°C	At 30°C
Tenth normal	0.3379	0.3376	0.3371
Normal	0.2860	0.2848	0.2835
Saturated	0.2496	0.2459	0.2420

pared than either of the others, and because it possesses certain other advantages over the other forms, the saturated type of calomel electrode is the one generally used in practice. When assembled according to definite specifications this electrode is both convenient and dependable. The absolute potential of a normal calomel electrode at 25°C is 0.560 volt. The mercury is + with respect to the KCl.

Two other special types of electrodes will be described in Sec. 96 in connection with a discussion of hydrogen-ion concentration.

93. The Primary Cell. We come now to a consideration of a source of electromotive force, viz., the primary or voltaic cell. This consists of two chemically dissimilar electrodes immersed in a suitable electrolyte or electrolytes. The term "primary" implies that any electrical energy delivered by the cell results from the direct chemical reaction between the electrolyte and the electrodes, one of the electrodes being used up as the reaction progresses.

Voltaic is a generic term, coming down to us from the time of Volta who made the first primary cells and who advanced the contact theory of potential. We have already considered (Sec. 91) the matter of electrode potentials. We may utilize the facts there noted in accounting for the emf developed by a primary cell.

Suppose we have a primary cell consisting of a piece of pure zinc immersed in a solution of zinc sulphate and a piece of pure copper immersed in a solution of copper sulphate, the two solutions being kept from rapid diffusion by means of a porous cup, as sketched in Fig. 106. This type of cell was originally designed by Professor Daniell in 1836 and bears his name. In the previous section the electrochemical reactions at the two electrodes were individually indicated. It was there pointed out that the zinc electrode was at a potential of 0.485 volt below that of the electrolyte and that the copper electrode was at a potential of 0.621 volt above its corresponding electrolyte. It therefore follows that **the potential difference between the two electrodes** will be the algebraic difference of these two potentials, or 1.106 volts; and this is the emf developed by the Daniell cell. In passing, it should be noted that in addition to the two sources of potential difference cited above there is, in the case of the Daniell and other similar cells, a third source of potential difference. If the ions of the two electrolytes do not move at the same velocity through their respective electrolytes there will exist a potential difference at the surface of contact between the two electrolytes. However, with the electrolytes commonly employed, this potential difference amounts to only a few millivolts and hence, for most practical purposes, may be neglected.

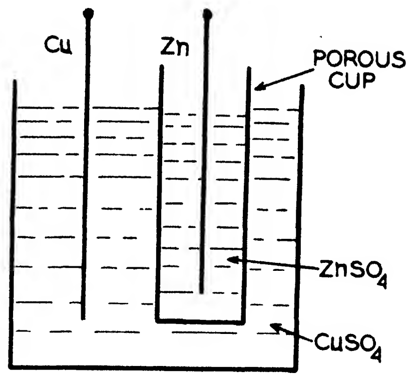


FIG. 106.—Daniell primary cell.

If the two electrodes of such a cell are electrically connected through a resistor, electrons will pass along the conductor from the negative terminal (zinc) to the positive electrode (copper), thereby destroying the electrochemical equilibrium that originally obtained at each electrode. This change results from the fact that the electronic migration, due to the emf at the terminals of the cell, tends to lessen the **electrostatic** force factor at each electrode, with the result that more of the zinc will go into solution and additional positive ions will go out of solution at the copper electrode. As previously pointed out, the terminal voltage of the cell under these circumstances will be less than the open-circuit potential, by an amount equal to the drop in potential through the cell itself. The Daniell cell possesses certain desirable characteristics that we shall examine shortly. For the present, its design and operation may be taken as representative of the primary cell as an electrochemical source of emf.

The primary cell may be thought of as an electrochemical transformer whereby chemical energy is converted into electrical energy.

94. Energy Relation in the Primary Cell. Again referring to the Daniell cell, let us examine the energy transformations which are involved in this typical case. Suppose that we connect the electrodes of our cell through a resistance of such a magnitude that 1 amp will flow in the circuit. Let this current flow for 1 sec. A quantity of electricity equal to 1 coulomb will, under those conditions, have passed completely around the circuit. By reference to our table of electrochemical equivalents (Sec. 89) it will be noted that the passage of 1 coulomb will require that 0.0003387 gm of zinc shall go into solution and that 0.0003294 gm of copper shall be liberated from the electrolyte. From thermochemical relations it is known that when 1 gm of zinc unites with sufficient SO_4 ions to be completely transformed into zinc sulphate ($ZnSO_4$) 1,630 calories of energy in the form of heat are liberated, this being what is known as the heat of formation per gram[1]. The electrochemical reaction at the zinc electrode would therefore involve an energy liberation given by the product of $1{,}630 \times 0.0003387$ or 0.553 calorie. Since one calorie is equivalent to 4.2×10^7 ergs, our reaction represents a liberation of 2.32×10^7 ergs.

Likewise, when 1 gm of copper unites with the acid radical to form copper sulphate ($CuSO_4$), 881 calories of heat are liberated. Therefore the amount of energy involved in **the rejection of the copper from the solution** would be given by $881 \times 0.0003294 \times 4.2 \times 10^7 = 1.22 \times 10^7$ ergs. The difference between the energy liberated at the zinc electrode and the energy required to deposit the copper at the other electrode would be the energy available for driving the current around the circuit, or

$$(2.32 \times 10^7) - (1.22 \times 10^7) = 1.1 \times 10^7 \text{ ergs.}$$

In general, the work done in transferring a charge around the circuit would be given by $Q \times V$ joules, or $Q \times V \times 10^7$ ergs. In this case, $Q \times V \times 10^7 = 1.1 \times 10^7$. In our example we made $Q = 1$ coulomb; hence

$$V = 1.1 \text{ volts} = \text{emf.}$$

As previously pointed out, the known emf of the Daniell cell is approximately 1.1 volts.

[1] The heat of formation per gram = heat of formation per gram equivalent per chemical equivalent. For $ZnSO_4$ the heat equivalent would be

$$248{,}000/(65.4 + 32 + 16 \times 4) = 1{,}630.$$

For a more extended discussion of heat of formation, see any standard work on chemistry.

The foregoing discussion is based on the assumption that the energy liberated as a result of the electrochemical reactions at the electrodes was completely converted into electrical energy. The close agreement of the computed and actual values indicates that such an assumption was warranted, at least in this particular case. There are, however, some cases in which all of the energy liberated by the reactions involved is **not** converted into electrical energy. In certain instances some of the liberated energy manifests itself directly in the form of heat. Cells of this type become warmer during operation. In such cases **the emf decreases as the temperature of the cell increases.** In certain other cases energy in the form of heat is abstracted from the components of the cell, and serves to augment the total energy available for maintaining the current in the circuit. Such a cell becomes colder when in operation and **its emf will show an increase with rise in temperature.** It is thus evident that some cells have what is known as a temperature coefficient, *i.e.*, their emf is a function of their temperature.

By the applications of the principles of thermodynamics to the reversible cell it is possible to work out a relation giving the emf of a cell in terms of the heat involved in the electrochemical reaction and the temperature change.

The term "reversible," in this connection, requires some explanation. Cells may be classified as **reversible** and **nonreversible** on the basis of the thermodynamic relations involved. These terms may be made clear by considering our typical cell, the Daniell unit.

Suppose we apply to the terminals of such a cell an external and opposite emf **slightly less** in value than the emf of the cell being considered. Under these conditions the cell will cause a small current to pass through the external agent and through the cell itself. Zinc will be dissolved from the negative electrode and copper will be deposited on the positive electrode. Suppose now that the externally applied emf be made **larger** than that of the cell being studied. The original process will now be reversed. Zinc will be deposited, and copper will go into solution. In short, the cell is restored to its original condition. If we neglect one or two very slight losses in this process such a cell may be considered to be **reversible.** If our two electrodes had been immersed in a single electrolyte such as hydrochloric acid, gas would have evolved when the cell was delivering current and the energy thus dissipated **could not have been recovered,** hence the process would have been **nonreversible.** It should be noted that the Daniell cell, and other similar units, are reversible **only when an infinitely small current is passing through the electrolyte.**

There is an important theorem, applying to reversible cells, that serves to express the emf in terms of the heat absorbed or given out by a

cell during the time when it is functioning as a source of electrical energy. This useful relation may be developed by taking a reversible cell through a series of electrochemical thermodynamic steps corresponding to Carnot's cycle.

A well-known theorem in thermodynamics is to the effect that in any reversible cycle operating between the absolute temperatures of T and T', the **ratio** of the thermal energy utilized to the **total** heat taken from the source equals the ratio of the difference between the two limiting absolute temperatures to the absolute temperature of the source. Algebraically, the above statement takes the form

$$\frac{H - H'}{H} = \frac{T - T'}{T}. \tag{i}$$

As the first step in the cycle, let the cell be held at the absolute temperature T, while it is allowed to liberate a **very small** quantity of electricity q, according to the chemical reaction

$$Zn + CuSO_4 \rightarrow Cu + ZnSO_4.$$

This will correspond to the line ab in Fig. 107, and constitutes an **isother-**

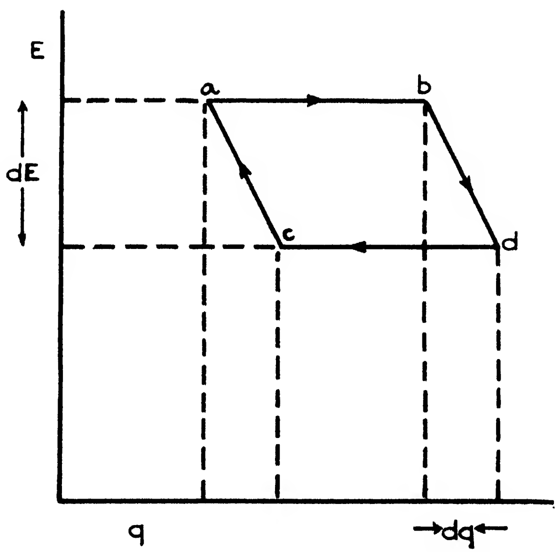


FIG. 107.—Illustrating the thermodynamics of a reversible cell.

mal change. The energy released during this process will be given by the expression

$$qE = H + H', \tag{ii}$$

where H is the net heat of formation, H' the external heat absorbed in order to hold the temperature constant, and E the emf of the cell.

The second thermodynamic step consists in thermally insulating the cell so that we may carry out an **adiabatic** process while allowing an additional small quantity of electricity to be drawn from the cell. Under these circumstances the cell will either heat or cool. Let us assume, for instance, that it cools, and that the temperature drops by an amount dT. This change is represented by the line *bd* on the diagram. At the end of this step the temperature of the cell will be $T - dT$.

The third step in the cycle consists in bringing about an **isothermal** change. To accomplish this we apply an opposing emf whose value is slightly greater than the normal emf of the cell. Under these conditions current will flow through the cell in a reverse direction, according to the chemical equation

$$Cu + ZnSO_4 \rightarrow Zn + CuSO_4.$$

Graphically this step will be represented by the line *dc*. During this step **work is done on the cell.**

The final step consists in passing a small additional quantity of electricity through the cell, still in the reverse direction, while the unit is thermally insulated, thus bringing about a final **adiabatic** changs. During this final step the temperature will rise to its initial value T, theu completing the cycle.

From elementary thermodynamic theory, we know that the area *abcd* in the diagram would represent the useful electrical work done by the cell during the above-indicated cycle. This area will be given by $q \times dE$, which product represents, then, the useful or net work done during the cycle. We may now substitute in (i) and arrive at the relation

$$q\frac{dE}{h} = \frac{dT}{T}.$$

Rearranging terms we have

$$h = qT\frac{dE}{dT}, \tag{iii}$$

where h represents the heat which must be supplied by, or given up to, the surroundings if there is to be no change in the temperature of the cell while it is delivering q units of electricity in the direction of its own emf. The term h corresponds to H' of (ii). The term dE/dT is an expression for the **temperature coefficient** of the emf. If dE/dT is **positive,** the emf of the cell will increase with increase of temperature; and heat, other than that supplied by the chemical reaction, must be supplied to the cell if it is to operate isothermally. In other words, $E > H$. Therefore,

when operating by itself, such a cell would cool when delivering electrical energy. If and when dE/dT is **negative** the emf will fall with rising temperature, which means that $E < H$. In such a case excess thermal energy is liberated by the chemical reaction, and the cell becomes warmer as it operates.

If now we combine (ii) and (iii) there results

$$qE = H + qT\frac{dE}{dT},$$

which leads to

$$E = \frac{H}{q} + T\frac{dE}{dT}.$$

If q is taken to be equal to a faraday, H will represent the heat supplied by the chemical reactions taking place in the cell when 1 gm equivalent of material is transformed. With this understanding the last equation may be written

$$E = H + T\frac{dE}{dT}. \tag{128}$$

If the temperature coefficient dE/dT is zero, the emf is determined solely by the heat of reaction per 96,494 coulombs of electricity delivered by the cell. In other words, when $dE/dT = 0$ the energy liberated by the chemical reactions in the cell is just sufficient to account for the emf of the cell, *i.e.*, $E = H$. Computed on the thermodynamic basis, the emf of the Daniell cell is 1.112 volts, which is in fairly close agreement with the value obtained on the assumption that its temperature coefficient is zero. The Daniell cell, because of the fact that it is reversible and because it has a low temperature coefficient, was used for many years as a standard cell.

As indicated above, the theorem embodied in Eq. (128) is an important relation. This expression was independently arrived at by Helmholtz and by Gibbs and is frequently referred to as the Gibbs-Helmholtz equation.

95. Concentration Cells. If a cell be constructed of two like electrodes immersed in a common electrolyte of uniform concentration, the potential difference between the electrolyte and the electrodes will be equal and oppositely directed, with the result that the effective emf of the cell will be zero. But we have already seen that electrode potential is a function of the ionic concentration. This is to be expected because of the fact that, as we have noted (Sec. 91), osmotic pressure is one of the factors that has a bearing on the electroequilibrium conditions in a primary cell; and osmotic pressure is known to be directly proportional to the concentration of the dissolved substance.

Since electrode potential does depend upon ionic concentration it is possible to assemble a cell having like electrodes and a common electrolyte which will show a terminal emf which is not zero. This is accomplished by arranging to have the electrolyte more concentrated in the region of one electrode than it is about the other. Such a cell is known as a **concentration cell.** The plan of a cell of this type may be represented thus:

Copper	copper sulphate (concentrated)	copper sulphate (dilute)	copper

The emf of such a cell is of small magnitude, being of the order of a few millivolts when the ionic concentration bears a ratio of 1:10. When such a cell is delivering energy the electrode in contact with the more dilute portion of the electrolyte goes into solution and metallic ions will be deposited on the other electrode. The latter will therefore be **positive** and the former **negative.**

The emf of a concentration cell is a function of the absolute temperature, as well as of the ratio of concentration. As a practical source of emf the concentration cell is of little value at present, but the fundamental principle involved does find wide application, as we shall see in the next section.

96. Measurement of Hydrogen-ion Concentration. In the field of chemistry the acidity or the alkalinity[1] of a solution is expressed in terms of the concentration of the hydrogen ions present in the solution. From one's study of elementary chemistry it will be recalled that the concentration of any ion is expressed numerically in terms of a normal solution, and that a **normal solution contains 1 gm of the ions per liter.** Such a solution would be expressed as $1N$. In our present discussion we are

[1] The connotation of the terms "acidity" and "alkalinity" is clearly indicated by the following excerpt from a work titled "Theoretical and Experimental Physical Chemistry" by Crocker and Matthews. "The characteristic acidic properties of an acid in solution depend upon the presence of hydrogen ions; and the alkaline character of a soluble base, in the same way, is dependent upon the presence of hydroxyl ions in solution. When these ions are present in equal proportions the solution is neutral. Thus water ionizes into equal proportions of hydrogen ions and hydroxyl ions, and hence, it shows no acidic or alkalinic reaction, if quite pure. When the hydroxyl ions are in excess of the hydrogen ions, then the solution is alkaline. When the hydrogen ions are in excess of the hydroxyl ions, the solution is acid; and the degree of acidity or alkalinity of the solution is directly proportional to the concentrations, respectively, of the hydrogen ions or of the hydroxyl ions. The determination of the degrees of dissociation of the acids or bases in solution, therefore, provides us with a means of comparing the relative strengths of acids and of bases." (The foregoing excerpt is reproduced by permission of the The Macmillan Company, publishers.)

interested in the hydrogen ionic content. Taking hydrochloric acid as an example, and assuming for the sake of simplicity that the atomic mass of hydrogen is unity, a 0.001 N solution of the acid would contain 0.001 gm of H ions per liter; its hydrogen-ion concentration would, accordingly, be $0.001N$ or $1 \times 10^{-3}N$. In most solutions the concentration of hydrogen ions is found to be a very small fraction of a gram per liter. It will thus be obvious that one will be dealing with cumbersome common fractions or with a large number of ciphers if the decimal form is used. To avoid this a logarithmic method of expressing ionic concentration is commonly employed. That such a plan is justified will be evident when we examine the theory upon which the measurement of ionic concentration is based. On a logarithmic basis, the ionic concentration could, in the case cited above, be indicated thus:

$$\begin{aligned} \log [H^+] &= -3 \\ -\log [H^+] &= 3 \\ \log \left[\frac{1}{H^+}\right] &= 3 \\ pH &= 3. \end{aligned}$$

The last expression appearing in the above list is the form most commonly used.

We next proceed to examine the experimental method whereby the pH value of a given solution may be determined. The technique involves the expression [Eq. (127)] giving the potential of an electrode in terms of electrolytic solution pressure and osmotic pressure. It has already been pointed out that the potential of the hydrogen electrode is a function of the hydrogen-ion concentration of the electrolyte with which the electrode is in contact. If Eq. (127) can be so modified that it will involve ionic concentration, one would then be able to determine the pH value of a given solution by measuring the potential between a hydrogen electrode and a second known electrode, both of which are in contact with the solution in question. Equation (127) is

$$E = \frac{RT}{nF} \log_e \frac{P_1}{P_2},$$

where P_1 is the solution pressure and P_2 the osmotic pressure. We know that osmotic pressure is proportional to ionic concentration; hence, the above expression may be written

$$E = \frac{RT}{nF} \log_e \frac{P_1}{kC}$$

where C is the hydrogen-ion concentration, and k a proportionality constant. But

$$\begin{aligned}\log_e \frac{P_1}{kC} &= \log_e P_1 - \log_e kC \\ &= \log_e P_1 - (\log_e k + \log_e C) \\ &= \log_e P_1 - \log_e k - \log_e C.\end{aligned}$$

Since, by previous definition, $E = 0$ when $P_1 = 1$ atm and $C_H = 1$ gm molecule/liter, it follows that

$$\log_e k = 0.$$

For the present case we have, then, that

$$\log_e \frac{P_1}{kC} = \frac{1}{C_H},$$

where C_H represents the hydrogen-ion concentration. Our relation thus becomes

$$E = \frac{RT}{nF} \log_e \frac{1}{C_H}.$$

Changing to common logarithms, and inserting numerical values for R and F, the above equation reduces to

$$E_H = 0.0001983T \log_{10} \frac{1}{C_H}, \tag{129}$$

where E_H is the potential of the hydrogen electrode referred to a normal hydrogen electrode.

In order to secure data from which to evaluate C_H, use is made of the assembly shown in Fig. 108. A hydrogen electrode and a calomel electrode are brought into contact with the solution whose pH value it is desired to determine. The potential difference between the mercury and the hydrogen electrodes is then measured by means of a potentiometer. The observed emf will be the sum of the two electrode potentials. Therefore, E (observed) $= E_{Hg} + E_H = E_{Hg} + 0.0001983T \times \log_{10} \frac{1}{C_H}$. Solving for the C_H term we get

$$\log_{10} \frac{1}{C_H} = \frac{E \text{ (observed)} - E_{Hg}}{0.0001983T} = pH. \tag{130}$$

And thus the hydrogen-ion concentration of any solution can be quickly and accurately determined.

As already pointed out, a $0.001N$ solution would have a pH value of 3. On the same basis the pH of a $0.01N$ solution would be 2; and for a $0.1N$ solution it would be 1. Pure water has a concentration of 10^{-7} gm of OH

ion per liter at 22°C. Such water is considered to be chemically neutral; hence, one may say that the pH of a neutral solution is 7. From the above examples it will be evident that **increasing** acidity is indicated by **decreasing** pH values; conversely, it may also be added that increasing alkalinity is shown by increasing pH values.

In some pH-determination work it is found more convenient to use one of two special electrodes in place of the hydrogen unit. This is particularly true in connection with industrial applications.

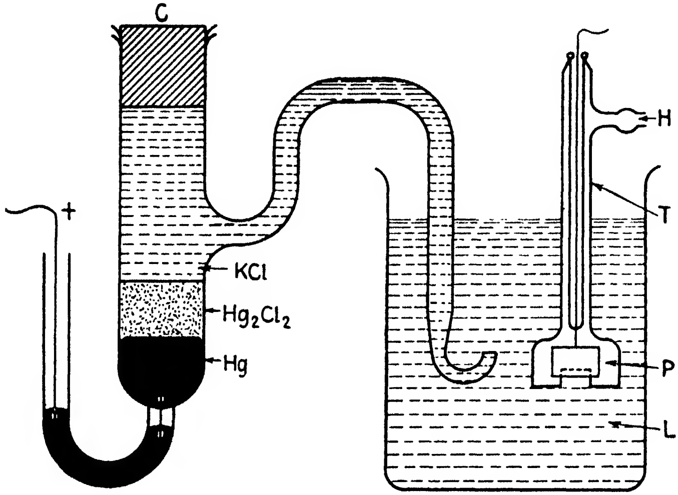


Fig. 108.—Hydrogen-ion determination.

In 1921, Biilmann[1] introduced what is known as the **quinhydrone electrode.** This electrode usually consists of a gold wire wrapped around the end of a glass tube into which one end of the wire is sealed. By means of the usual mercury-wire combination within the tube, connection is made between the exposed gold wire and the potentiometer. When such an electrode is immersed in a solution to which has been added a small amount of **quinhydrone** it is found that the potential of the electrode is directly related to the hydrogen-ion concentration of the solution and its temperature. Quinhydrone is a compound consisting of quinone ($C_6H_4O_2$) and hydroquinone [$C_6H_4(OH)_2$] combined in equimolecular proportions. When dissolved the compound dissociates into its two constituents, one of which ionizes. When equilibrium obtains the condition is indicated thus:

$$\text{Quinone} + 2H^+ + 2e \rightleftarrows \text{hydroquinone}.$$

[1] Biilmann, E., *Ann. Chemie*, **15**, 109 (1921).

If and when the ratio between the concentration of the quinone to the concentration of hydroquinone is constant and approximately unity, the relation between hydrogen-ion concentration and electrode potential is linear, within limits. Since quinhydrone is only slightly soluble in acid, only a small quantity is required to bring about the condition specified above. The potential of a gold or platinum electrode in contact with a solution saturated with quinhydrone is given by the relation,[1]

$$E = \text{constant} + \frac{RT}{F} \log_e C_H.$$

The above equation may be changed to the form

$$pH = \frac{0.7177 - 0.00074t - E - V}{0.0001983T}, \qquad (131)$$

where t is the temperature of the solution and the electrodes, E the em reading of the potentiometer (potential difference between the calomel and gold electrode), and V the potential of the calomel electrode at the working temperature. In an acid solution the quinhydrone electrode is higher than that of the saturated calomel electrode, and hence must be connected to the positive terminal of the potentiometer. However, at a temperature of 25°C and a hydrogen-ion concentration (pH) of 7.64 the potential of the two electrodes becomes equal. If the pH value of the solution is higher than the value just indicated, the calomel electrode is higher than the gold electrode and hence the connections must be reversed. The quinhydrone electrode is not reliable when the pH value is above 8, but below this limit it is entirely dependable, and very convenient to use.

In recent years still another form of electrode has come into extensive use. Reference is made to the so-called **glass electrode.** In 1909 Haber discovered that a potential difference exists between a glass surface and a solution, and that this potential difference is a linear function of the hydrogen-ion concentration of the solution. This type of electrode consists of a thin-walled bulb of low-melting-point glass (Corning 015) in which is placed a solution containing hydrogen ions (buffer solution) plus a small amount of quinhydrone into which dips a platinum wire that serves as a connection. The glass electrode, thus constructed, is placed in contact with the solution under investigation and the potential difference between this electrode and a reference electrode (usually a calomel unit) is measured by means of a special potentiometer network. The glass electrode is chemically inert and can be used in the determination of

[1] See "Theoretical and Physical Chemistry" by Crocker and Matthews for a discussion of this equation.

pH values up to 9. Within this range the glass electrode is 0.352 volt positive with respect to a hydrogen electrode in the same solution. In practice, however, it is customary to calibrate the glass electrode by means of solutions the hydrogen-ion concentration of which is accurately known.

Since a glass wall forms an element in the electrical circuit, the resistance of a cell, one component of which is a glass electrode, is very high, being of the order of 10^7 to 10^8 ohms. As implied above, this necessitates

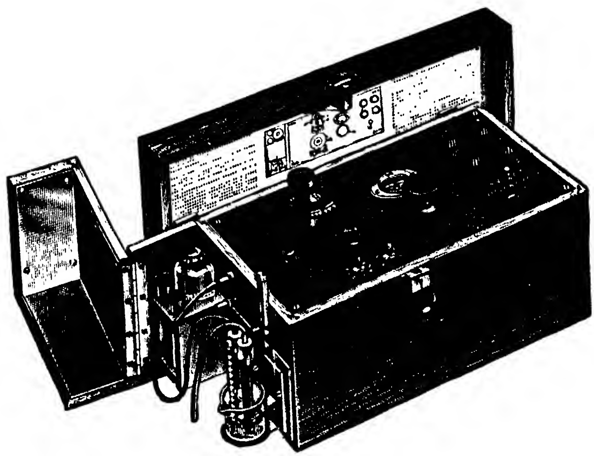


Fig. 109.—Commercial form of hydrogen-ion determination assembly, using a glass electrode. (*Leeds & Northrup Co.*)

the use of special potential-measuring facilities. To avoid polarizing the cell made up of the glass electrode and the reference electrode, and also to make it possible to utilize a rugged type of galvanometer, recourse is had to an electronic amplifier (Sec. 237), usually of the single-tube type. By means of a suitable bridge and a potentiometer that incorporates a tube amplifier, it is possible to construct a compact portable assembly that can be made to read *pH* values directly. A commercial unit of this character is shown in Fig. 109. A detailed discussion of the circuits involved is beyond the scope of this book.

Because of the wide use to which a knowledge of pH values is put in both research and industry, it is highly important that all students of physics, chemistry, and biology acquire a working knowledge of the theory and technique of making pH determinations. For useful information regarding the theory and applications of the glass electrode, the reader is referred to a recently published monograph entitled "The Glass Electrode," by M. Cole, published by John Wiley & Sons.

97. Polarization. Returning again to the problems directly connected with the operation of primary cells, we encounter a fundamental phenomenon that manifests itself whenever certain types of primary cells are employed as a source of emf.

We have already shown (Sec. 94) that the passage of current through a reversible cell does not change the chemical nature of the electrodes or the electrolyte. For instance, it has been pointed out that, having a cell consisting of two copper electrodes immersed in copper sulphate, the application of the slightest emf will cause some current to pass through the cell. If, however, we have a combination consisting of carbon electrodes immersed in zinc sulphate and apply a definite external emf to the terminals of such a cell, it will be found that after a brief interval the current through the cell will diminish in value and, unless the applied emf exceeds 2 volts, the current will entirely cease. In the short time during which the current is passing through the cell a thin layer of zinc will be deposited from the solution on the negative terminal (cathode), and therefore one of our original carbon electrodes has been electrochemically transformed into another type of electrode. We now have a cell made up of carbon and zinc electrodes, with zinc sulphate as the electrolyte. A test will show that the emf of this combination is about 2 volts, and that the direction of the emf is opposite to that of the original applied emf, the carbon now being positive with respect to the zinc-plated rod.

The counter emf thus established is known as the **emf of polarization** and the phenomenon connected with the production of this opposing emf is known as **polarization.**

It is also possible to bring about polarization under somewhat different circumstances. If, instead of using zinc sulphate as our electrolyte in the cell just described, we had employed dilute sulphuric acid, decomposition of the solution would also have occured; but in this case hydrogen gas would have appeared at one of the electrodes (negative) and oxygen at the other. A counter emf would be manifest in this case also, having a value of about 1.7 volts. In this instance we have produced two dissimilar electrodes, one consisting of an extremely thin layer of hydrogen gas about the original cathode, and the other being composed of molecules of oxygen clinging to the original anode. Thus we have established what amounts to a cell in which the electrodes are gaseous. In other words, polarization has occurred; and in order to force a current through either this or the former cell, we must apply an emf greater than the emf of polarization.

When the process of polarization is brought about under circumstances similar to those outlined in the first case it may be made use of in important ways, as we shall see shortly. The type of polarization exemplified

in the last instance, however, becomes a troublesome phenomenon, particularly in connection with the use of primary cells. It does not alter the essentials of the case if the gas is liberated **as a result of the passage of the current through the cell due to its own emf** rather than as a result of an externally applied emf. The liberation of hydrogen as a result of the operation of the cell itself, and its appearance at one of the electrodes, constitutes one of the serious defects of primary cells. Not only is a counter emf set up, but the existence of the layer of gas molecules introduces a high resistance into the internal circuit of the cell. As a result, the effective emf of the cell is decidedly reduced. In order to prevent polarization a number of plans have been devised to keep the hydrogen from reaching the positive electrode. It is chiefly because of these various depolarization schemes that we have the several types of existing primary cells. In Sec. 99 we shall consider this matter further.

98. Local Action. Because of its position in the electrochemical series, (Sec. 92), and also because of its relative cheapness, zinc is most commonly employed as the material for one of the electrodes in primary cells. Due to the impurities frequently present, what is known as **local action** takes place. A bit of impurity, iron for instance, forms a tiny local primary cell with the surrounding zinc and electrolyte with the result that a local emf is developed which in turn gives rise to a locally circulating current. Thus the electrode tends to disintegrate even when the cell is not in use.

To prevent this, recourse is had to what is known as amalgamation. This consists in bringing mercury mechanically or chemically[1] into contact with the surface of the zinc, and thus forming a zinc-mercury amalgam. The impurities do not unite with the mercury and are covered up by the zinc-mercury amalgam formed on the surface of the electrode. As the zinc is dissolved the mercury remains and unites with the remaining zinc, the impurities falling to the bottom of the cell.

Local action may also be caused in some cells by concentration effects. We have seen (Sec. 95) that an emf may result from a difference in concentration between two points in an electrolyte. Near the surface of the electrolyte the density will tend to be less than in the body of the solution. This gives rise to a local emf and the consequent local wasting of the electrode at a point near the surface of the solution.

99. Examples of Primary Cells. In designing a primary cell there are several important ends to be attained, among which may be mentioned (1) highest possible emf, (2) rapid and complete depolarization, (3)

[1] The most satisfactory method of amalgamating battery zinc is by immersion for a few minutes in a solution made by dissolving mercury in "aqua regia." Convenient proportions are mercury, 15 cc; nitric acid, 170 cc; hydrochloric acid, 625 cc.

absence of local action, (4) low internal resistance, and (5) low cost of materials. To the foregoing characteristics might be added the feature of portability which is important for certain classes of service. How some, at least, of these desirable features have been attained is illustrated by several of the primary cells in current use. It may not be out of place, however, to say that the "last word" in the design of primary cells probably has not been said. It is quite possible that further research in this field would result in a substantial monetary return.

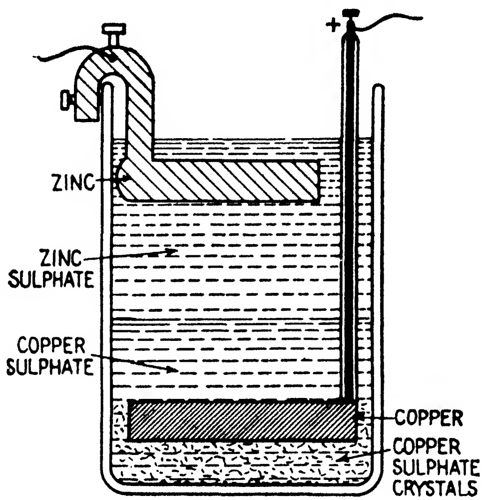


FIG. 110.—Gravity form of primary cell.

We have already examined the Daniell cell. While its emf is quite constant and depolarization is complete, its **internal resistance** is high—which fact limits its usefulness in cases where currents of any magnitude are desired. A cell that has a somewhat lower internal resistance than the original Daniell unit, but which is essentially of the same type, is commonly known as the **gravity cell** (Fig. 110). This cell consists of a negative electrode composed of zinc, commonly in the form of a "crowfoot"; a number of copper strips riveted together serve as the positive terminal. A saturated solution of copper sulphate surrounds the copper electrode, and the zinc electrode is immersed in a solution of zinc sulphate. Partial separation of the two electrolytic solutions is maintained as a result of the difference in density of the two electrolytes, hence the name.

Depolarization is rapid and complete in both this and the Daniell cell, and is accomplished as a result of the presence of the two electrolytes. Zinc passes into solution thus,

$$Zn + H_2SO_4 \rightarrow ZnSO_4 + 2H^+ + 2e,$$

Simultaneously, at the positive electrode, copper is deposited as given by the equation

$$CuSO_4 \rightarrow Cu^{++} + SO_2^{--}.$$

Where the two solutions are in contact the H ions and the SO_4 ions unite to form sulphuric acid thus,

$$2H^+ + SO_4^{--} \rightarrow H_2SO_4.$$

It will thus be seen that hydrogen does not reach the copper electrode and hence polarization does not occur. The gravity cell gives an emf of about 1.08 volts, but its internal resistance is comparatively high, being of the

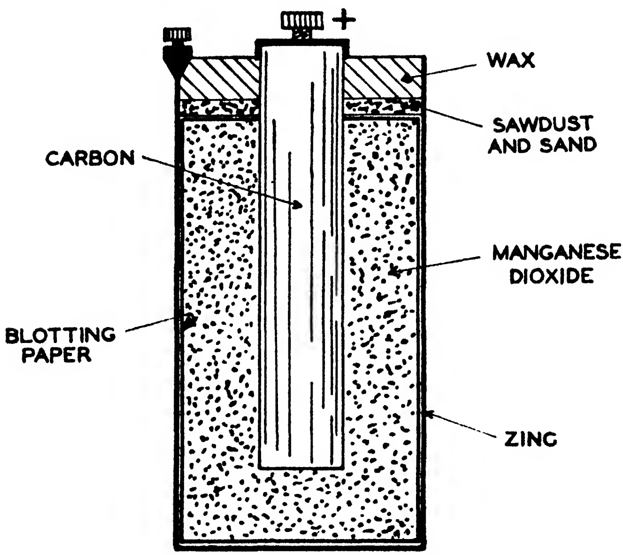


FIG. 111.—Dry cell.

order of 1 ohm. Hence, the cell can deliver only a relatively small current even on short circuit. Because of the fact that polarization is absent, this cell may be used in closed-circuit work. It has been extensively employed in telegraphic circuits but is not widely used at present.

The modern **dry cell** is an outgrowth of a cell originally known as the Leclanché cell. The original unit consisted of a zinc and a carbon electrode in a single electrolyte, ammonium chloride (sal ammoniac). In the Leclanché cell the positive electrode (carbon) was inclosed in a porous cup in which were packed manganese dioxide and granular carbon. Diffusion of the solution took place through the porous cup and its contents. In the modern portable form of this cell the porous cup is dispensed with and the zinc electrode forms the container (see Fig. 111).

The positive electrode (carbon rod) forms the center of the unit. This is surrounded by a mixture of manganese dioxide, granular carbon, and zinc chloride. This, in turn, is surrounded by a layer of absorbent material such as sawdust, and a layer of blotting paper next to the zinc container. The sawdust and paper are saturated with a solution of ammonium chloride. The top of the cell is sealed to prevent evaporation.

When current is passing through the cell we have, as a result of dissociation,

$$NH_4Cl \rightarrow NH_3 + H^+ + Cl^-.$$

At the negative electrode,

$$Zn + 2Cl^{--} \rightarrow ZnCl_2 + 2e.$$

Simultaneously, at the carbon electrode the hydrogen undergoes oxidation thus,

$$2MnO_2 + 2H^+ \rightarrow Mn_2O_3 + H_2O - 2e.$$

It will be noted that the hydrogen unites with oxygen to form water. Under the circumstances this reaction takes place slowly; hence, if the circuit is closed for any appreciable length of time, excess hydrogen will accumulate and polarization will occur. This type of cell is therefore not adapted for closed-circuit work. Notwithstanding this limitation, it is very extensively used in the United States. The emf of the dry cell is about 1.5 volts and the internal resistance of the standard No. 6 size varies from 0.05 to 0.1 ohm.

Another form of single-electrolyte cell finds application in signal-circuit work, particularly in connection with the operation of railroad block signals. This cell utilizes two zinc plates connected in parallel as the negative terminal, and a block of compressed copper oxide held in a copper band serves as a positive electrode (Fig. 112). The electrolyte is a concentrated solution of sodium hydroxide (caustic soda). A layer of heavy oil is placed on top of the electrolyte, thus preventing evaporation and the "creeping" of the solution. The cell is ruggedly built and may be left for long periods without attention. It was originally devised by Dr. Lalande, but the modern form is largely due to Edison, and is now known as the **Edison primary cell.**

When this cell is delivering energy, we have, at the negative electrode,

$$Zn + 2NaOH \rightarrow ZnNa_2O_2 + 2H^+ + 2e.$$

At the copper electrode the hydrogen, in reducing the copper oxide, is itself oxidized thus,

$$2H^+ + CuO \rightarrow H_2O + Cu - 2e.$$

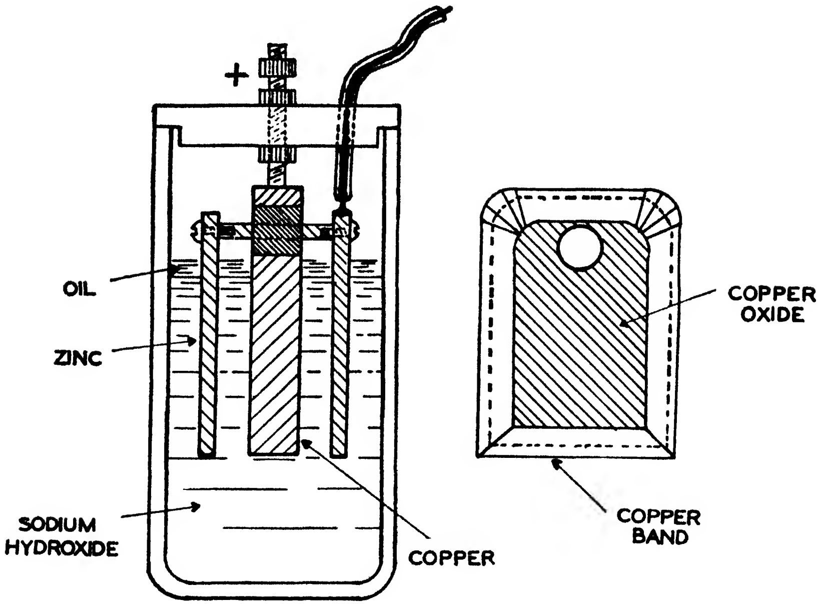


FIG. 112.—Edison primary cell.

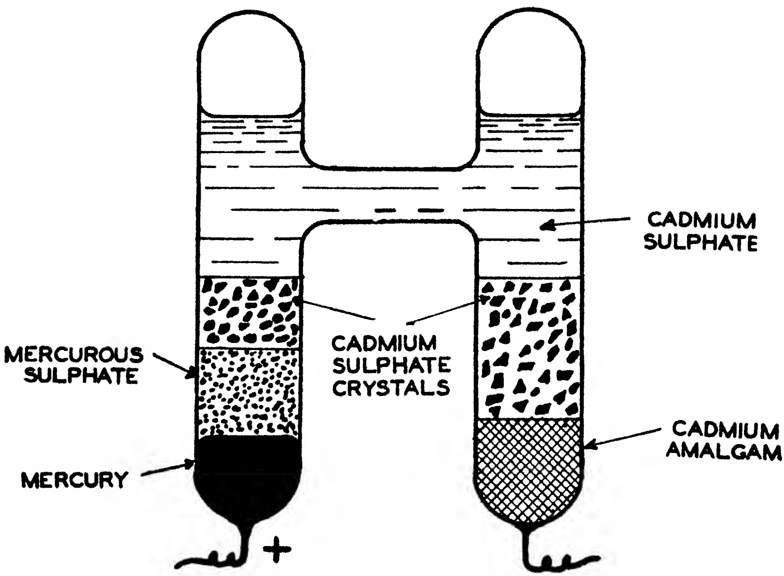


FIG. 113.—Weston standard cell.

The last reaction takes place rapidly, and hence polarization does not occur. Local action is also absent. The emf of the cell is about 0.75 volt. Owing to the large area of the electrodes, and the short distance of separation, the internal resistance is low, varying from 0.02 to 0.1 ohm, depending on the size of the unit. The cell is therefore capable of delivering large currents. The commercial form of the Edison primary cell is made in sizes ranging from 100 to 600 ampere-hours[1], and the elements of a given cell are so designed that they all become exhausted at about the same time, thus facilitating renewal.

The primary cell giving the most constant emf is known as the **Weston standard cell.** Figure 113 shows the general plan of the cell assembly. Mercury in contact with saturated mercurous sulphate forms the positive electrode, while cadmium amalgam, in contact with saturated cadmium sulphate, serves as the negative element. This cell was designed by Dr. Edward Weston and has, because of its reliability, become the standard of emf throughout the world. The International Electrical Conference (London, 1908) set up specifications for the construction of the Weston cell; and when the unit is assembled according to those instructions the emf of different cells will not differ from one another by more than two or three parts in 100,000. The temperature correction formula recommended by the London Conference is

$$E_t = 1.01830 - 0.0000406(t - 20°C) - 0.00000095(t - 20°C)^2 + 0.00000001(t - 20°C)^3,$$

where E_t is the emf of the cell in international volts at any temperature t on the centigrade scale. Commonly the first two terms of this relation are used. In using a standard cell it is important to keep in mind that no current greater than 0.0001 amp should ever be allowed to pass through such a cell, and then only for a few seconds. If any appreciable current is drawn from the cell a certain amount of polarization will occur and thus cause a reduction in the cell's emf, thereby destroying the cell's value as a standardizing unit. The Weston cell is one of the two or three standard units upon which all electrical measurements are based.

100. Secondary Cells (Storage Batteries). In our discussion of the theory of primary cells (Sec. 93) attention was directed to the fact that such a cell consists essentially of two electrodes of **unlike material,** and that, during the use of the cell, at least one of these electrodes was gradually **converted by electrochemical action into a different substance.** It was further noted that, in the case of certain primary cells (the reversi-

[1] Batteries are rated in terms of ampere-hours. For instance, a 150-ampere-hour cell will deliver 10 amp for 15 hr. All batteries have a **normal discharge rate.** If this rate is exceeded the cell will not, in general, deliver its rated output.

ble type), the cell, after use, might be restored to its original condition by sending a current through the unit in the reverse direction. The Daniell cell was found to be of this type. Obviously, then, such a cell could be utilized for the conversion of electrical energy into chemical potential energy **and for the reconversion of chemical potential energy into electrical energy.** When used for such a purpose, the cell would be called **a secondary cell,** or **accumulator.** While such a process is theoretically possible in the case of any reversible cell, certain practical considerations make it advisable to utilize a special combination of electrodes and electrolytes in the construction of secondary or storage cells.

From our previous study of the theory of cells it is evident that in order to produce a cell that will have a maximum emf we should select two elements which are far apart in the electrochemical series. It happens that if a **compound** consists of a metal and some very electronegative element, such as oxygen, the compound (in this case the oxide) will be electronegative with respect to the metal; and the greater the percentage of oxygen in the compound the more electronegative will it be. It is an interesting fact that the oxide of a comparatively cheap metal, viz., lead, is highly electronegative with respect to lead itself. Furthermore, lead and lead oxide (PbO_2) in sulphuric acid form a reversible cell.

In his work on the polarization of metals in electrolysis, M. Gaston Planté in 1860 observed that an electrolytic cell showed a reverse emf after the applied emf was disconnected, and that this phenomenon was particularly marked when lead served as the electrodes. This was the beginning of the development of the lead accumulator.

To assist in understanding the theory of the secondary cell let us examine what happened in Planté's original experiment. If we immerse two pieces of clean lead in dilute sulphuric acid and connect the terminals to a source of emf exceeding 2.2 volts electrolysis will take place. Hydrogen will be liberated at the negative electrode (cathode) and oxygen at the positive terminal (anode). The metallic lead will be attacked by the oxygen, forming a coating of brown lead oxide (PbO_2) on the anode. The hydrogen will not react with lead. We began our hypothetical experiment with two like plates. As a result of electrolysis, one of these electrodes, in part at least, was changed to an entirely different substance, so that we now have a lead electrode and a lead dioxide electrode. In other words, by expending energy in the form of the electrical current we have manufactured a **new electrode.**

If now we disconnect the source of outside current and test our cell it will be found that the electrode coated with lead dioxide is positive with respect to the metallic lead plate—the terminal emf being about 2.2 volts. It will also be found that if we connect the terminals of our cell to some

electrical device, such as a door bell, current will flow for several minutes, gradually decreasing in value.

If we examine our cell after having used it as a secondary source of electrical energy, it will be found that electrochemical reaction has resulted in the changing of the lead oxide back to metallic lead. The cycle of operations has thus been completed. During the first stage of our experiment, commonly but erroneously called "charging," electrical energy was converted into **chemical potential energy.** During the second stage of the process this chemical potential energy was made use of to

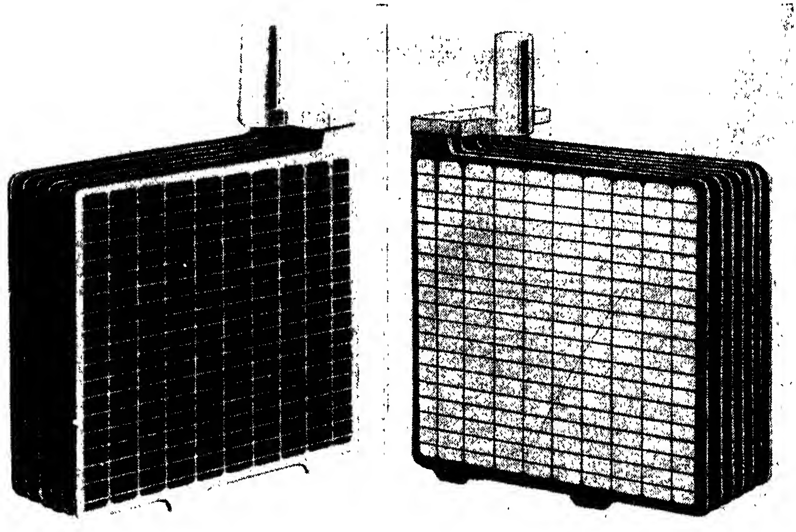


Fig. 114.—Positive and negative elements of a Faure type of secondary cell. Positive components at the left; negative at the right. (*The Electric Storage Battery Co.*)

produce an electric current exactly as is the case in a primary cell. After the cell was "charged" it contained no more electrons than before the process started.

In the simple experiment just described, only a comparatively small amount of oxide was formed and hence but little energy was "stored." Planté found that it was possible to increase materially the amount of oxide formed by a given charging current if the plates were first put through a process known as **forming.** This consisted of a series of reversals—charging first in one direction, discharging, allowing to stand for a time and then charging in the reverse direction. By this process a superficial layer of the electrode is transformed into "active material" which results in a greatly increased efficiency. The process of forming,

however, is quite expensive, and as a result the Planté type of plate has to a great extent been replaced by one made by a more rapid and less expensive procedure.

M. Camille Faure introduced a type of electrode for secondary cells which is known as the **Faure or pasted-plate electrode.** This consists of a grid of lead (Fig. 114) into the interstices of which is forced a paste consisting of lead oxide and dilute sulphuric acid. Red lead (Pb_3O_4) is used in the positive grid and litharge (PbO) in the negative plate. At least a part of the mixture on each plate is chemically changed to lead sulphate ($PbSO_4$). As a result of the use of this "pasting" process a part of the work of forming the plates is done chemically, and thus the time of preparing the plates for use is greatly lessened and the manufacturing cost materially reduced.

The Faure type of electrode is lighter in weight but is not so rugged as the Planté plate. Owing to its lighter weight and relative cheapness the Faure type of plate is widely used for portable units. It is also used to some extent in large fixed plants.

While the exact electrochemical reaction that occurs in a lead secondary cell is somewhat uncertain, the following equations probably represent what takes place. The dissociation of the electrolyte gives

$$H_2SO_4 \rightarrow H_2^{+} + SO_4^{--}.$$

$$\text{During charge} \ldots\ldots \begin{cases} \text{At the positive plate,} \\ PbSO_4 + SO_4^{--} + 2H_2O \rightarrow PbO_2 + 2H_2SO_4 + 2e \\ \text{At the negative plate,} \\ PbSO_4 + 2H^{+} \rightarrow Pb + H_2SO_4 - 2e \end{cases}$$

$$\text{During discharge} \ldots \begin{cases} \text{At the positive plate,} \\ PbO_2 + H_2SO_4 + 2H^{+} \rightarrow PbSO_4 + 2H_2O - 2e \\ \text{At the negative plate,} \\ Pb + SO_4^{--} \rightarrow PbSO_4 + 2e \end{cases}$$

These reactions show that during the charging process lead dioxide is formed on the positive grid and spongy lead on the negative plate. The amount of sulphuric acid in the electrolyte also increases, and hence the specific gravity of the solution rises. While the cell is delivering energy in the form of electric current both the lead and the lead dioxide revert to lead sulphate, thus completing the electrochemical cycle. During these electrochemical reactions the so-called "active material" in the plates undergoes considerable expansion and contraction. As the cell is repeatedly charged and discharged nonreversible reactions also tend to occur. Both of these factors operate to limit the life of the battery. The positive plates disintegrate more rapidly than the negative grids.

The **ampere-hour capacity** of a secondary cell is proportional to the amount of active material available in the plates. To secure large capacity, cells are usually made of a series of positive and negative plates, with like grids connected in parallel. Such an arrangement not only serves to augment the capacity, but also serves to reduce the internal resistance, which, in most cases, is only a few hundredths of an ohm. If a cell is to be used under conditions requiring a heavy discharge current

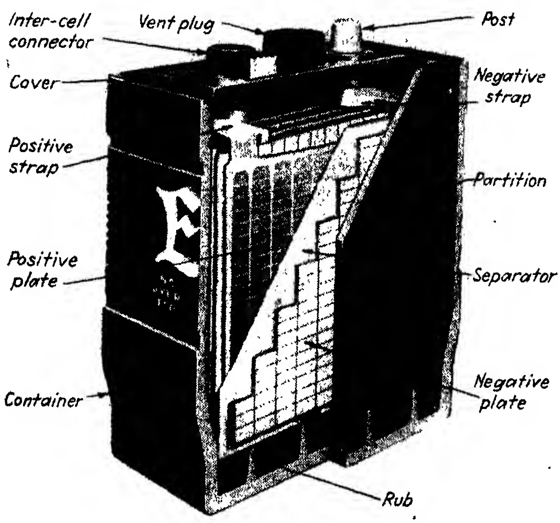


FIG. 115. Cutaway view of a commercial lead storage cell. (*The Electric Storage Battery Co.*)

the plates are made comparatively thin, thus allowing the electrolyte freer access to the active material, and hence producing a more rapid reaction. A sectionalized view of a modern lead storage cell of the portable type is shown in Fig. 115. Figure 116 gives the charge and discharge characteristics of a lead cell.

In speaking of the ampere-hour capacity of a secondary cell it should be borne in mind that the rate of discharge is an important factor in this connection. Any given cell is designed to be charged and discharged at a certain predetermined rate known as the normal charge and discharge rate. The ampere-hour rating of a cell is based on its normal rate of discharge.

A freshly charged secondary cell shows a terminal emf of about 2.2 volts which becomes less on discharge, and should never be carried below 1.8 volts. Further discharge brings about certain nonreversible reactions

in the active material, thereby more or less permanently reducing the capacity of the cell. In practice it is found advisable to take the density or specific gravity of the electrolyte as an index of the condition of the battery. The concentration of the electrolyte used in a given cell depends somewhat on the type of plates used and the character of the service for which the battery is designed. In the case of stationary batteries the specific gravity of a fully charged cell will be of the order of 1.23 and when completely discharged will show a reading of about 1.15. The electrolyte of cells used in connection with automobiles and for other similar purposes has a specific gravity of 1.27 to 1.30 when charged and 1.15 to 1.18 when discharged.

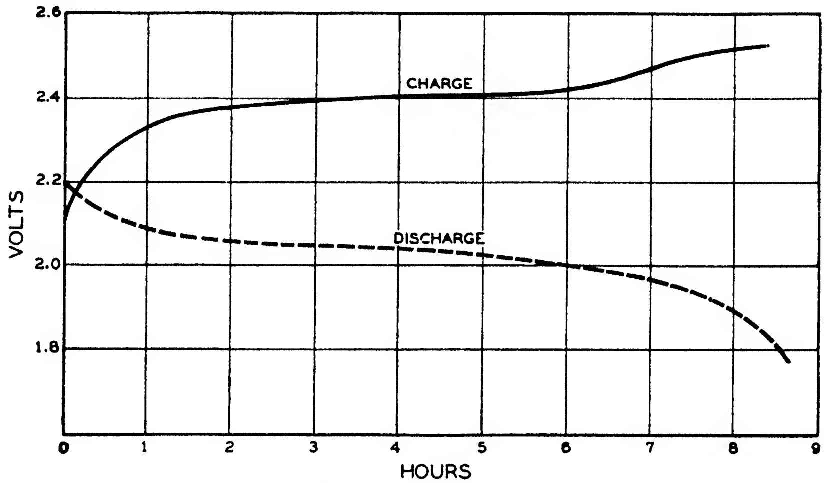


FIG. 116.—Representative charge and discharge curves in the case of a lead cell.

By the **efficiency** of a storage battery is meant the ratio of the watt-hour output to the watt-hour input. On this basis a good secondary cell has an efficiency of about 75 per cent. When given proper care and adapted to the class of service it is called upon to render, the life of a good accumulator should be not less than 200 cycles of charge and discharge.

While the lead secondary cell has a comparatively high efficiency and is extensively used for many purposes, it has, nevertheless, certain inherent disadvantages, chief among which are its weight, its tendency to lose capacity with use, the more or less undesirable character of the electrolyte (a strong acid), and the necessity of careful supervision at all times. Many attempts have been made to produce an accumulator unit in which these features would be absent, or at least be present to a lesser degree. The **Edison storage battery** has, to some extent, accomplished this end.

In the secondary cell developed by Mr. Edison we have a unit that, in some respects, is radically different from the lead cell. The active material of the positive plate in the Edison cell is nickel hydrate [$Ni(OH)_2$], and

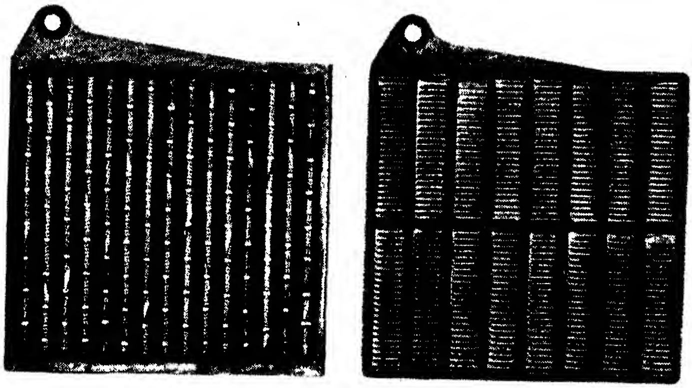


FIG. 117.—Positive and negative plates of an Edison secondary cell. Positive unit on the left.

that of the negative plate, iron oxide (FeO). The electrolyte is a 21 per cent solution of potassium hydrate (KOH), to which has been added a small amount of lithium hydrate.

The mechanical construction of the plates, and the cell as a whole, is somewhat unusual. Figures 117 and 118 show the detailed construction and plan of the cell assembly. The positive grid consists of a group of small perforated nickel-steel tubes rigidly fastened to a nickeled-steel frame. These tubes are packed with alternate layers of nickel hydrate and pure nickel flake, the latter constituent being introduced for the purpose of increasing the electrical conductivity of the active metal. The negative grid is made up of a series of perforated nickeled-steel pockets into which is packed the iron oxide and to which is added a small amount of mercury, the latter element serving to increase the conductivity. The metallic pockets are rigidly fastened to a nickeled-steel frame. The entire assembly of positive and negative grids is solidly

FIG. 118.—Cutaway view of an Edison secondary cell.

bolted together and supported in a housing that consists of a nickel-plated steel container. It will thus be seen that the Edison cell is an extremely robust unit and therefore mechanically adapted to severe types of service.

The exact chemical reactions that take place in the Edison cell have not as yet been definitely determined. However, it is probable that the following represent about what takes place.

By dissociation we have

$$KOH \rightarrow K^+ + OH^-.$$

$$\text{During charge}\ldots\ldots\ldots \left\{\begin{array}{l} \text{At the positive grid,} \\ 2Ni(OH)_2 + 2OH^- \rightarrow 2Ni(OH)_3 + 2e \\ \text{At the negative grid,} \\ FeO + 2K^+ + H_2O \rightarrow Fe + 2KOH - 2e \end{array}\right.$$

$$\text{During discharge} \left\{\begin{array}{l} \text{At the positive grid,} \\ 2Ni(OH)_3 + 2K^+ \rightarrow 2Ni(OH)_2 + 2KOH - 2e \\ \text{At the negative grid,} \\ Fe^{++} + 2OH^- \rightarrow FeO + H_2O + 2e \end{array}\right.$$

From these reactions it is evident that the electrolyte remains unchanged during charge and discharge; hence, there is no change in the specific gravity of the solution during a cycle of operation. The only function performed by the electrolyte is to serve as a medium for the transfer of the hydroxyl ion from one plate to the other. It would appear that the nickel-iron-hydrate combination, as arranged in the Edison cell, makes possible **a group of completely reversible reactions.** In this respect this cell differs from the corresponding lead unit in which the reactions are not completely reversible, and become less so as the age of the battery increases.

The capacity of the nickel-iron cell tends to increase somewhat after it is put into commission. The effective life of the Edison portable units is at least three times that of the lead cell and in some instances much greater than this. On the basis of watt-hours per pound, the Edison cell is considerably lighter than the lead unit.

The electrical characteristics of the Edison cell are also different from those of the lead unit. Its terminal emf is about 1.4 volts when fully charged, falling to approximately 1 volt when discharged. Thus the average emf is about 1.2. The internal resistance of the cell is somewhat higher than that of its rival, resulting in an energy efficiency of something like 60 per cent. The cell is not damaged by short-circuiting, can be completely discharged without injury, and may be left unused, either charged or discharged, for an indefinite time. Because of the fact that

there is no change in the specific gravity of the electrolyte, the hydrometer test would give no indication of the state of charge or discharge; hence, the condition of the cell is judged by the terminal voltage. Because of the nature of the materials entering into the construction of the Edison cell its initial cost is about five times that of the lead unit, but when length of battery life is considered the cost is approximately the same for both types.

For detailed discussion of the theory, manufacture, care, and application of storage batteries, the reader is referred to a treatise by G. W. Vinal, entitled "Storage Batteries."

101. Grouping of Cells. In order to secure an emf greater than that given by one primary or secondary cell a number of units may be connected in **series,** as shown in Fig. 119*a*.

In this case the total emf would be given by nE where E represents the emf of a single cell (assuming all of the cells to have like values of emf), and n the number of cells connected in series. If we assume that each cell in the series has an internal resistance of r, the current that would flow through the circuit, if the switch were closed, would be given by the relation

$$I = \frac{nE}{R + nr}, \tag{132}$$

where R is the value of the external or **load resistance.** As previously pointed out, when current is flowing the terminal voltage of the cell assembly will be less than the total emf by an amount equal to the IR drop through the cells.

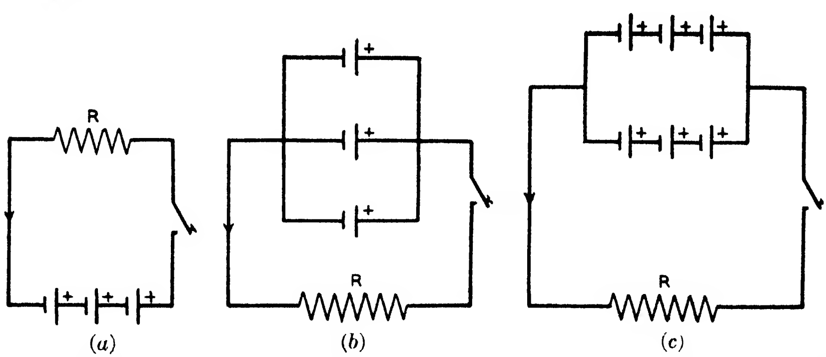


FIG. 119.—Grouping of battery cells: (*a*) in series, (*b*) in parallel, (*c*) in series-parallel.

In the event that the external resistance R becomes vanishingly small the current would be given by the familiar $I = E/r$.

If it is desired to use a current greater than can be safely delivered by a single cell, a parallel grouping may be arranged as shown in Fig. 119*b*.

In this case the total internal resistance would be $1/n$th that of a single cell, or r/n. Thus the current through the load resistance R would be

$$I = \frac{E}{(r/n) + R}. \tag{133}$$

If R is very large compared to r the current value would approach E/R as a limit.

If a high total emf and a high current value are both desired a series-parallel arrangement of cells will attain such an end, as indicated in Fig. 119c. Under these circumstances the current would be given by the relation

$$I = \frac{nE}{(nr/m) + R}, \tag{134}$$

where n is the number of cells in each series group and m the number of rows in parallel. In the illustration shown $n = 3$ and $m = 2$.

The question may arise in practice as to what arrangement of cells will give the greatest current through a given external (load) resistance when each cell has a known internal resistance. If we divide both numerator and denominator of the fraction in Eq. (133) by n we get

$$I = \frac{E}{(r/m) + (R/n)}.$$

Now, since E is constant, I will be at a maximum when the denominator has a minimum value. But the denominator is the sum of two terms whose product is a constant. It therefore follows that the denominator will have a minimum value when

$$\frac{R}{n} = \frac{r}{m},$$

or when

$$R = \frac{nr}{m},$$

where nr/m is the internal resistance of the battery. It is therefore apparent that maximum current will obtain when the total resistance of the battery is made equal to the load (external) resistance.

It might be found useful to have available a relation by means of which one might express the maximum current in terms of the total number of cells and the resistances involved, although in practice it is seldom possible to arrange the factors so that this condition holds.

By multiplying the last equation above by n, and rearranging terms, we get

$$n^2r = nmR = NR,$$

where N represents the total number of cells making up the series-parallel combination. From this we find that

$$n = \sqrt{\frac{NR}{r}}.$$

We have seen that, if the load and total internal resistances are equal,

$$R = \frac{nr}{m}.$$

If, then, we substitute these values for n and R in Eq. (133) there results

$$I_{max} = \frac{E}{2}\sqrt{\frac{N}{Rr}}. \qquad (135)$$

But it is to be remembered that the relation holds only when the external (load) resistance is equal to the total internal battery resistance.

PROBLEMS

1. A saturated calomel electrode was used to determine the potential of a given electrode when immersed in a solution of one of its own salts. A potentiometer showed a reading of 0.908 volt when the temperature of the solution was 25°C. What would be the potential of the electrode under test, and what was the probable nature of the electrode material?

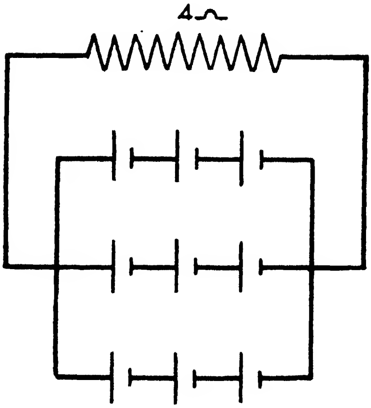
2. Assume a primary cell is made up of zinc and mercury electrodes in contact with a zinc sulphate and a mercurous sulphate solution, respectively. Compute from the electrochemical relations the emf that would be developed.

3. In a certain pH determination, when using a hydrogen and a saturated calomel electrode at 25°C, the potentiometer gave a reading of 0.54 volt. What was the hydrogen-ion concentration of the solution?

4. The hydrogen-ion concentration of another solution was investigated by means of the potentiometer, a calomel electrode, and a quinhydrone electrode, all at a temperature of 25°C. If the potentiometer showed a reading of 0.1 volt, what is the pH value?

5. A bank of 55 lead storage cells connected in series supplies a load resistor of 5 ohms. On the assumption that the internal resistance of each cell is 0.1 ohm, compare the energy loss, in watts, in the external and the internal parts of the circuit. What would be the value of the current if the battery were short-circuited?

6. A number of cells are grouped as shown in the accompanying sketch. If the emf of each cell is 1.5 and the individual cell resistance is 0.1 ohm, what will be the value of the current through the resistor?



7. With the above-indicated arrangement of cells, what would be the magnitude of the current if the load resistance were made equal to the total internal resistance of the battery?

8. Given 10 cells each having a resistance of 0.1 ohm, determine the nearest practical grouping that will give the maximum current through a load resistance of 5 ohms.

9. A two-wire telegraph line is 30 miles in length. The wire is No. 10 copper having a resistance of 1.02 ohms/1,000 ft. Four telegraph relays are to be operated in series on the line, the resistance of the winding of each relay being 200 ohms. The instruments require 25 ma for satisfactory operation. What is the smallest number of gravity cells that can successfully operate the relays? The emf per cell is 1.08 volts, and the internal resistance per cell is 3 ohms.

10. Suppose that 10 Edison-Lalande cells connected in series are used to operate a railway semaphore signal. On the assumption that the signaling mechanism is operated on an average of 30 min per day, how often must the zinc electrodes be renewed? Each cell has an emf of 0.7 volt, and the operating current is 10 amp. Two zinc plates are connected in parallel in each cell, and the plates weigh 625 gm each.

CHAPTER XIV

THERMOELECTRIC PHENOMENA

102. The Seebeck Effect. In the preceding chapter we studied the means whereby an emf may be produced as a result of chemical reactions. In 1822, J. Seebeck announced to the Berlin Academy of Sciences the discovery of an entirely new method of producing an electric current. Seebeck arranged a circuit of two metallic bars one of which was copper and the other bismuth, the two pieces of metal being soldered together at their ends, as illustrated in Fig. 120. A magnetic needle was placed as shown. When one of the junctions was warmed Seebeck found that the magnetic needle was deflected, and in such a direction as to indicate that a current flowed across the hot junction (as he put it) from the bismuth to the copper. The **electronic** current passed from the copper to the bismuth, showing that the hot end of the latter element is at the higher potential. Further investigation showed that **the magnitude of the effect is a function of the difference in temperature between the two junctions** and also that various combinations of metals give different results. It thus became evident that thermal energy may be utilized to establish an emf, and thus it can be directly converted into electrical energy. A pair of metals so arranged that one junction of the pair may be maintained at a different temperature than the other and thus be used to develop an emf has come to be known as a **thermoelement** or **thermocouple.**

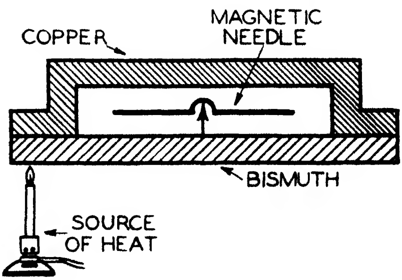


FIG. 120.—Generation of a thermoelectric current.

On the basis of his observations Seebeck arranged a number of the metals in a series so that when any pair of the metals is employed as a thermocouple the conventional current flows across the hot junction from the one occurring earlier in the series to the one appearing later in the list. The electron current would be the reverse of this. Seebeck's list included, among others,

Bismuth	Mercury	Silver
Nickel	Lead	Zinc
Cobalt	Tin	Tungsten
Palladium	Chromium	Cadmium
Platinum	Molybdenum	Iron
Uranium	Rhodium	Arsenic
Copper	Iridium	Antimony
Manganese	Gold	Tellurium
Titanium		

At best, the magnitude of the emf developed by a thermocouple is very small. For instance, using an antimony-bismuth pair, the emf developed per degree centigrade difference in temperature between 0 and 100° is 0.000057 volt. Hence if one of the junctions of such a couple were maintained at 0° and the other at 100° the emf produced would be less than 6 millivolts.

Notwithstanding the low emf produced by thermal means, it was hoped, immediately following Seebeck's discovery, that the thermocouple might be utilized for the production of electrical energy directly from heat on a practical scale. Indeed, Clamond designed a thermoelectric battery consisting of a large number of heavy metal bars so arranged that the emf developed would be additive when similar junctions of the entire series were simultaneously heated. Employing 120 pairs and heating the junctions by means of a gas flame, he secured an emf of 8 volts; the efficiency, however, was extremely low, being less than 1 per cent.

Though the thermojunction method of producing current for power purposes has not proved feasible, nevertheless this device has found an important field of usefulness in connection with temperature and radiation measurements. As a result of the work of the Italian physicist, Meloni, and several American investigators, among whom may be mentioned A. H. Pfund and W. W. Coblentz, the sensibility of the thermocouple has been greatly increased, and has thus become a research tool of great importance. We shall discuss certain uses of this in a later section.

103. Laws of Addition of Thermal Emf. Three simple but important laws in connection with the use of thermocouples have been experimentally established. In using a thermocouple, it is obviously necessary to introduce some current- or potential-indicating device into the circuit in series with the metals forming the original junction. This naturally involves the presence of additional metallic contacts in the circuit which will, in general, constitute thermojunctions, and which may therefore develop thermal emf. The question arises as to how these additional junctions will affect the emf developed in the circuit as a whole. Experi-

ment has shown that the introduction of one or more pieces of metal into the circuit does not change the total emf, **provided the junctions thus introduced are maintained at the same temperature as that of the point in the circuit where they are inserted.** This fact is known as the **law of intermediate metals.**

For instance, in the circuit shown in Fig. 121 the presence of the copper connecting wires will not alter the emf developed in the circuit if the temperature of the points x and x' is maintained at the same value, this value being what it was before the connecting wires were inserted. It is therefore evident that the two metals M_1 and M_2 forming the couple may be soldered together at one of the junction points without changing the emf developed by the couple.

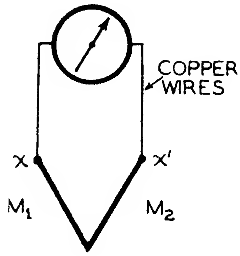


FIG. 121.—Effect of connections on the emf developed by a thermojunction.

The second law has to do with the relation that exists between the temperature and the emf developed, and is referred to as the **law of successive temperatures.** This law is to the effect that **for any given thermocouple the emf developed when the junctions of the couple are maintained at any two specified temperatures is equal to the sum of the emf that would be produced if the couple was operated between successive temperature steps throughout the original range.** Symbolically this fact may be represented thus:

$$E_{t_1}{}^{t_n} = E_{t_1}{}^{t_2} + E_{t_2}{}^{t_3} + \cdot \cdot \cdot + E_{t_{n-1}}{}^{t_n} \qquad (136)$$

where t_1, t_2, t_3, etc., are any successive temperature values between t_1 and t_n. We shall find this relation useful in the next part of our discussion.

A third law, based on experimental findings, is to the effect that if we have one thermojunction consisting of two metals, A and B, and a second junction consisting of metals B and C, the two junctions being in series, the over-all emf will be equal to the emf developed if and when the metals A and C function as a thermojunction, and if both junctions are operated at the same temperature. An application of this property of the thermojunction will be found in connection with thermoelectric diagrams shown in the following section.

104. Thermoelectric Diagrams. Before proceeding to a description of practical applications of the thermocouple we may well examine the relation which obtains between the temperature factor and the emf developed.

If we set up a thermocouple circuit as sketched in Fig. 122, and apply heat as shown, thereby gradually raising the temperature of the oil-

immersed junction, the deflection of the galvanometer will increase for a time. If the heating is continued beyond a certain temperature, however, the current will gradually decrease to zero and will, in fact, reverse

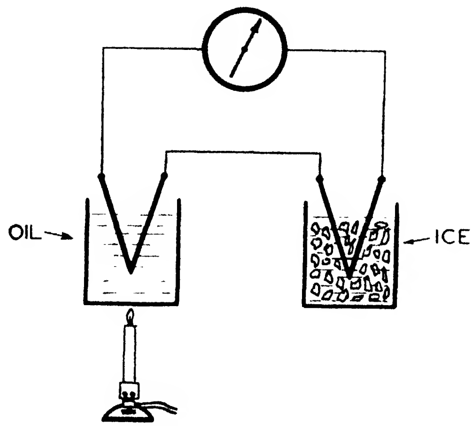


FIG. 122.—Experimental arrangement for determining the relation between thermal emf and temperature.

in direction if the heating is continued far enough. If we make a graph from the data obtained by such an experiment a curve of the form shown in Fig. 123*a* will result. It is found that a similar curve results for other thermocouple combinations.

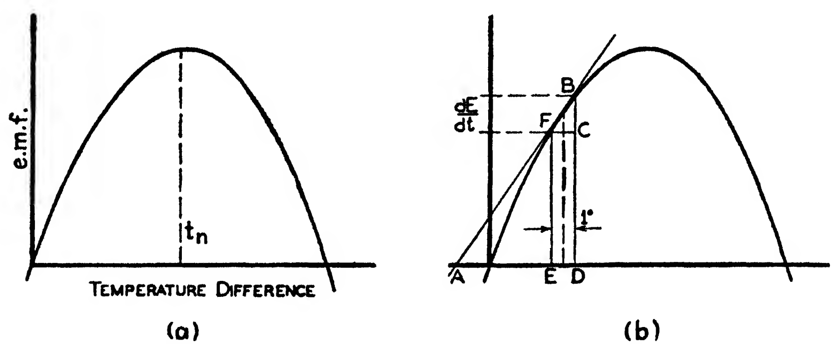


FIG. 123.—Relation between emf and temperature in the case of a thermojunction.

An examination of the curve will disclose the fact that it is parabolic in form (axes parallel to emf axis), and consequently may be represented by the equation $y = ax + bx^2$. In our case this relation becomes

$$E_0{}^t = at + bt^2, \tag{i}$$

where E_0' represents the emf; t the temperature of the hot junction, the other being at 0°; and a and b are constants that depend on the nature of the particular pair of metals used. If we differentiate (i) with respect to t, we get

$$\frac{dE}{dt} = a + 2bt, \tag{ii}$$

in which dE/dt will represent the slope of the curve (Fig. 123b), and hence the rate of change of emf with temperature. If we make the temperature interval 1°, represented by ED or FC in Fig. 123b, the corresponding change in emf will be represented by BC. Under these conditions dE/dt is known as the **thermoelectric power** of the particular couple.

If in (ii) we made dE/dt equal to zero, the resulting expression will give the temperature t_n at which the emf reaches a maximum. That temperature is given by the relation $t_n = -a/2b$, and is known as the **neutral point.** It is to be noted that the neutral point does not depend upon the temperature of the cold junction. If the temperature of the cold junction is higher than zero it is equivalent to moving the temperature axis upward in the graph diagram—the form of the curve is not changed.

Inspection of (i) shows that E is a two-valued function, and hence there are **two** temperatures of the hot junction which will give the same value for the emf developed. Thus, if a thermojunction is employed as a temperature-indicating device, it is important to know on which side of the neutral point one is operating, otherwise the reading will be ambiguous. A solution of the quadratic equation (i) gives

$$t = \frac{-a \pm \sqrt{a^2 + 4bE}}{2b}, \tag{iii}$$

thus confirming the above statement regarding the uncertainty of t unless t_n is known. In the above relation, if and when the quantity under the radical is equal to zero,

$$t = -\frac{a}{2b},$$

which, as shown above, is the neutral temperature.

Again referring to (iii), it is also evident that there are two temperature values for which E will be zero. One such value is zero and the other is $-a/b$. When $t = -a/b$ the temperature of the hot junction is as far above t_n as the neutral point is above the cold junction. When $(a^2 + 4bE)$ equals zero, it follows that $E = -a^2/4b$, which is the maximum positive value that the emf can attain in the case of that particular couple.

In practice, it is frequently convenient to know the value of the emf developed by some given pair of metals when operated between any two known temperatures. Professor Tait has shown that it is possible to set up a simple diagram by the aid of which one may readily compute the value of the emf produced in any given case.

Referring again to (ii), it will be noted that it is of the form

$$y = a + 2bx,$$

and that this represents a straight line. Furthermore, the tangent which the line in question makes with the X-axis is given by $2b$.

Suppose then, say in the case of lead and zinc, we plot the thermoelectric power, dE/dt, against temperature, as shown in Fig. 124. The

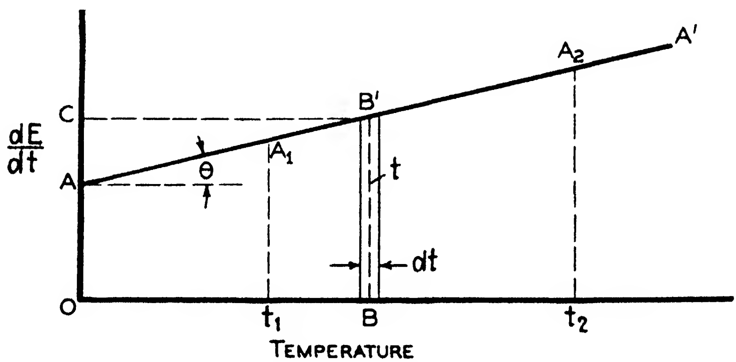


FIG. 124.—Temperature-emf relation in the case of a lead-zinc thermojunction.

straight line AA' will result. In this graph OA represents the constant a and $\tan \theta = 2b$. At a temperature represented by t, the median point of the narrow temperature range dt, the thermoelectric power will be represented by OC, the height of the elemental strip BB'. The area of this strip $= (dE/dt)dt = dE$. Therefore the elemental area BB' represents the small emf dE developed by the thermocouple when its two junctions are at the infinitesimal difference of the temperature represented by dt.

From the law of successive temperatures (Sec. 103) it follows that the **total** emf developed by the couple, when its junctions are maintained at the temperatures t_1 and t_2 respectively, will be represented by the area $A_1A_2t_2t_1$ in Fig. 124. Geometrically the area is equal to the product of one-half the sum of the two thermoelectric power ordinates and the distance representing the difference in temperature $t_1 - t_2$. It may therefore be said that **the emf developed by a thermocouple whose junctions are maintained at two different temperatures is numerically**

equal to the product of the thermoelectric power and the difference in the temperatures.

If now we plot on the same diagram a second thermoelectric power-temperature line for a pair of metals such as, say, lead and iron, we will have a line something like BB' in Fig. 125. The line AA' applies to our

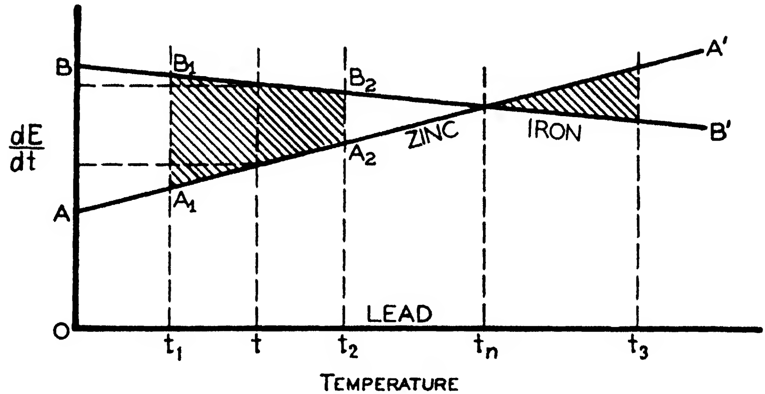


FIG. 125.—Thermoelectric power-temperature relation for two junctions having one element in common.

original pair, lead and zinc. From what has already been said, it will be evident that the emf developed by the lead-iron pair will be represented by the area $B_1B_2t_2t_1$. As a result of the third property of a thermojunction (Sec. 103), it follows that the emf developed by a junction composed of zinc and iron, when operated between the same temperatures (t_1 and t_2), will be represented by the area $B_1B_2A_2A_1$. From the geometry of the case the magnitude of this emf will be given by the product of the difference in temperature of the two junctions ($t_2 - t_1$) and the difference in the thermoelectric power when referred to lead at the mean temperature t of the two junctions. If one of the temperatures, say t_3, falls on the other side of the neutral point t_n, the emf developed in the circuit will be given by the **difference** in the two shaded areas.

It is thus evident that if some one metal be taken as a standard of reference we can extend our diagram to include as many metals as desired. For reasons to be pointed out later, lead is usually taken as the reference metal, though platinum is also sometimes utilized as a basis. Such a thermoelectric diagram showing the lines for a number of metals is given as Fig. 126. The figure is based on one given by Noll in "Wiedemann's Annalen," 53, p. 874.

An an example of the use of the thermoelectric diagram, suppose we were to maintain one junction of our zinc-iron combination at 50° and

the other at 150°. The mean temperature would therefore be 100°. From the diagram (Fig. 126) we see that the mean thermoelectric power for zinc and for iron at 100° is approximately 7 μv. The total emf that

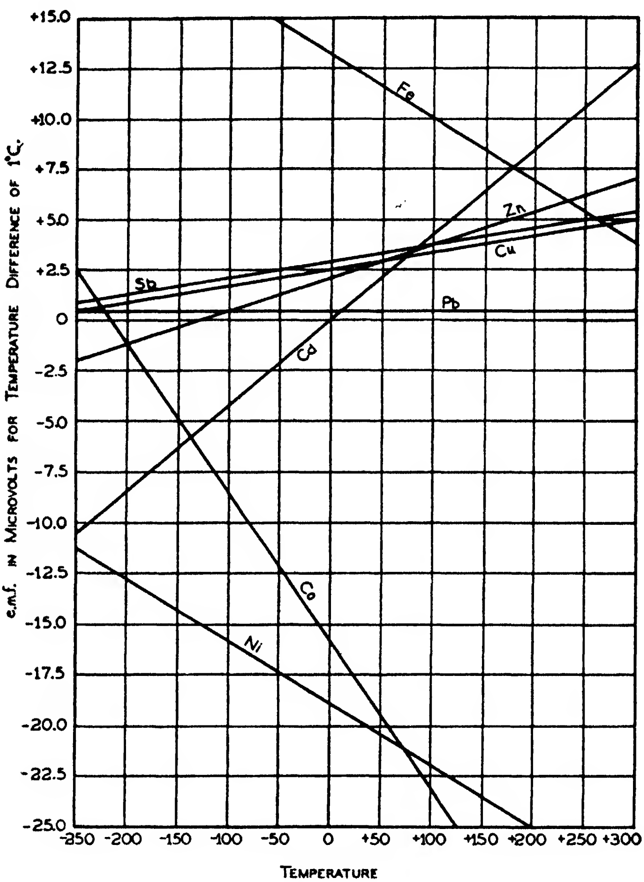


FIG. 126.—Thermoelectric diagram.

would be developed by the zinc-iron couple under the conditions named would be 7 × 100 = 700 microvolts.

The student will find a list of thermoelectric power values for various metals given in "Smithsonian Physical Tables," 7th ed., pp. 317 *ff.*

105. Peltier Effect. In 1834, 12 yr after Seebeck (Sec. 102) made his discovery, J. C. Peltier of Paris discovered a phenomenon which is the converse of the Seebeck effect. He demonstrated that an electric current when passed across a thermojunction will, under certain circumstances, cause an increase in temperature; and under certain other conditions a

lowering of the temperature. For instance, when Peltier passed a current across an antimony-copper junction from the former to the latter element, he secured a rise of 10° in temperature; when the current was reversed, the temperature was dininished 5°. An antimony-bismuth couple gave even more marked effects. In fact it is said that Lenz, of whom we shall hear later in another connection, was able to freeze water by the Peltier effect.

It is not difficult to account for the Peltier effect. Our discussion of the Seebeck effect (Sec. 102) led to the conclusions that the application of heat to a metal changes its electrical properties and that different elements are affected differently. (We are not referring here to changes in electrical resistance.) Furthermore the law of conservation of energy must hold in the case of thermocouples as well as in other connections.

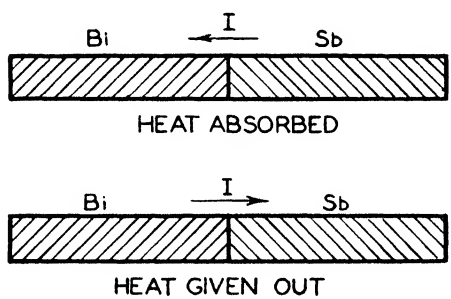


FIG. 127.—Peltier effect.

If we refer again to Seebeck's original experiment, as shown in Fig. 120, it is evident that an emf was developed at the hot junction, and that the emf caused a current to flow around the circuit. The energy represented by the current was supplied by the applied flame; in other words, **heat disappeared at that junction.** In our study of the primary cell (Sec. 93) we saw that, if no energy were dissipated, the process which gave rise to an emf was reversible. In the case now under discussion the process proves to be, in part at least, a reversible one. When a bismuth-antimony junction is heated the direction of the electronic current is as shown in Fig. 127. If the external heat source be removed and current from an outside source be passed across the junction, heat will be absorbed and the junction cooled. Conversely, if current from an outside source is sent across the junction in a direction opposite to the normal thermoelectric current, as indicated in Fig. 127, the current will be converted into heat and the heat will be liberated. These phenomena are entirely consonant with the law of the conservation of energy.

It can also be shown that the Peltier effect is in conformity with the laws of thermodynamics; but it is beyond the scope of this volume to enter into this aspect of the case. In passing, however, it may be noted that the heat absorbed by the thermoelectric current at the heated junction is given out again at the colder junction; thus a thermocouple acts as a heat engine.

In considering the Peltier effect, the student is cautioned not to confuse this phenomenon with the thermal effect due to the resistance encountered by the current in passing through all conductors. In the latter case the heat produced varies as I^2R, and is independent of the direction of the current. In the Peltier phenomenon the thermal effect varies as the first power of the current.

106. Thomson Effect. In 1856 Lord Kelvin (then Professor William Thomson) was led to believe, from theoretical considerations, that an effect similar to the Peltier phenomenon might be expected to obtain throughout the body of any conductor when the conductor was **unequally heated.** Experiment has shown that there is a potential difference between different parts of a conductor when these parts are not at the same temperature. In any given case the direction of the emf depends upon the relative temperature and the nature of the conductor. For instance, in the case of copper the emf is directed from the colder to the hotter parts. In the case of iron, palladium, cobalt, nickel, and platinum the reverse is true. Zinc, tin, silver, and several other metals act similarly to copper. Lead does not show an appreciable Thomson effect and **it is for this reason that lead is taken as the reference metal in the thermoelectric diagram.** In the case of those metals which behave as does copper the Thomson effect is said to be positive. In the conductors that behave similar to iron it is considered to be negative. On the thermoelectric diagram (Fig. 126) those metals having lines that slope downward to the right show a negative Thomson effect, and those that slope upward exhibit a positive Thomson effect.

107. Applications of Thermocouples. Shortly after Seebeck's disclosure of the elementary principles of the thermocouple an Italian physicist, Leopoldo Nobili, a professor in Florence, devised what is known as the **thermopile.** This consisted of small alternate bars of bismuth and antimony connected in series, the bars being suitably insulated from one another except at the ends which were soldered together. Thus there was formed a "battery" of thermocouples, the pairs being connected in series, as shown diagrammatically in Fig. 128. The junctions on one side of the "pile" were blackened in order to absorb completely any incident radiation.

Macedonio Melloni, a contemporary and fellow countryman of Nobili,

greatly improved the thermopile and utilized it extensively in connection with his classical researches in the field of thermal radiation.

In recent years the thermocouple has been still further improved and now serves as the essential part of several research and engineering devices of remarkable sensitiveness and wide utility. For example, multiple couples consisting of fine iron and constantan wires are now used for measuring the radiant energy in spectrographic lines. A single thermocouple made of exceedingly fine wires and directly connected to a high-sensitivity galvanometer is sometimes employed as a radiation detecting system in connection with temperature studies of distant stars.

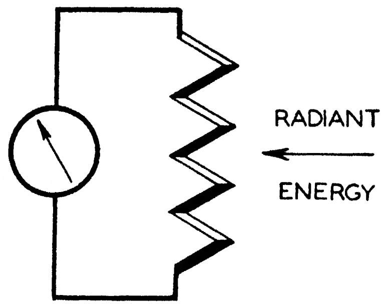


FIG. 128.—Thermojunctions in series.

A highly sensitive instrument of this general type, known as a **radiomicrometer,** was devised by Professor C. V. Boys, and used by him in the study of radiant energy. The essentials of this device are shown in Fig. 129. It consists of a delicate thermocouple and galvanometer combined in one instrument. A loop of silver wire W terminates in a bismuth-antimony pair J. The closed electrical circuit thus formed is suspended by a fine quartz fiber between the pole pieces, N and S, of a strong magnet. Any rotation of the movable system is indicated by a beam of light reflected from the mirror M to a suitably positioned scale. In order to absorb completely all incident radiation a small piece of blackened copper C is attached to the active junction. Because of the fact that a magnetic field produces certain effects on antimony and bismuth, and to avoid extraneous thermal effects, the thermocouple hangs in a thick-walled iron housing. An opening is provided in this inclosure through which the radiation is admitted to the thermojunction.

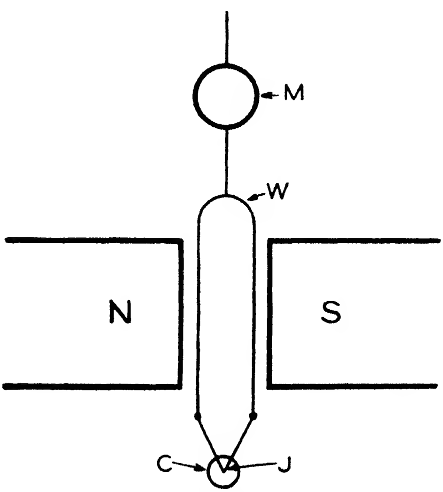


FIG. 129.—Boys' radiomicrometer.

The slightest heating of the lower junction will give rise to a current around the loop which, due to the strong magnetic field, will result in a

rotation of the movable system. Incredibly small quantities of radiant energy have been detected and measured by this instrument. It is said that it will give an appreciable deflection when actuated by the energy from a candle a quarter of a mile distant.

A modification of Boys' radiometer was made by W. Duddell in 1889, and is known as the **Duddell thermogalvanometer.** It is an instrument designed to measure small values of alternating current (Sec. 155). The only essential difference between the original Boys instrument and the Duddell modification consists in the manner in which heat is caused to reach the thermocouple. In the Duddell thermogalvanometer a small resistance element is placed directly beneath the thermocouple. The current to be measured is passed through the resistance and heats the adjacent thermojunction, thus causing a deflection of the movable system due to the current in the loop. Figure 130 indicates the essential components of the Duddell instrument. The instrument is made with several heater units, thus adapting it to various classes of service. This type of galvanometer is quick acting, nearly aperiodic, and has practically no capacitance or self-inductance (Sec. 133). Because of these characteristics it is particularly adaptable to the measurement of extremely small values of h-f alternating current. Currents as low as 20 microamperes may be measured. Using a suitable resistor the current generated by a sustained tone spoken into a common telephone receiver is sufficient to give a full-scale deflection.

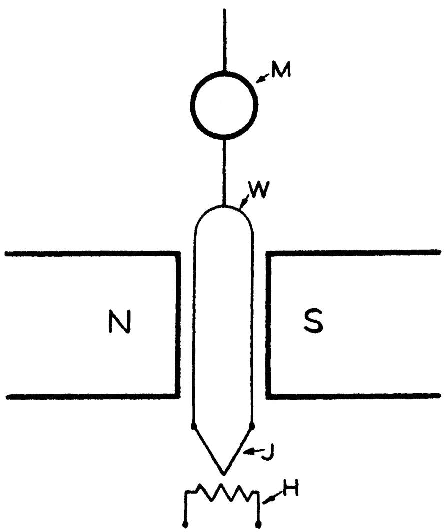


FIG. 130.—Duddell thermogalvanometer.

Portable and switchboard ammeters utilizing a thermocouple as a detecting agent, but operating on a somewhat different plan from the Duddell instruments, are now available for research and commercial processes. The general design on which these instruments are based is shown schematically in Fig. 131. The current to be measured is passed through a resistance H which serves as a heater. The active junction of the thermocouple J is in direct contact with the resistor. Connections from the terminals of the thermocouple are made to a suitable d-c indicating instrument, such as a sensitive galvanometer or a milliammeter.

By reducing the mass of the heater and thermojunction elements to a minimum, and inclosing these parts in a vacuum, the sensitivity of the organization is materially increased. Thermocouple ammeters may be accurately calibrated by means of direct current, and are suitable for use in the measurement of alternating currents of any frequency and wave form.[1]

While the thermocouple is used to a considerable extent in the measurement of the electric current, it is as a temperature and radiation-indicating agent that it finds its most extensive research and industrial application.

When one considers the possible use of the thermocouple as a temperature-indicating device, an interesting and important problem at once presents itself. On referring to Fig. 123*a* it will be seen that there are two possible temperature values corresponding to any given emf value. If, then, we take the emf developed by a couple as an index of temperature, the result will, in general, be ambiguous. In order to avoid this, a pair of metals must be selected that have a neutral point well above the maximum temperature it is desired to measure. Two metals that have nearly parallel thermoelectric lines (Fig. 126) would fulfill this requirement, and such a pair would have an emf-temperature curve which would be nearly straight. Fortunately there are several metals of this character available.

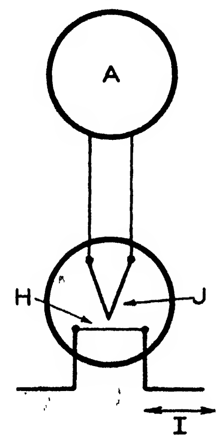


FIG. 131.—Thermocouple type of ammeter.

In selecting the elements from which to assemble a thermojunction, two factors must be considered: (1) the magnitude of the emf per degree that will be developed; and (2) the effect of extreme temperatures on the physical and electrical properties of the metals forming the couple. A vast amount of research has been carried out in this connection.

At present there are two general types of thermoelements employed and they are designated as **base-metal couples** and **noble-metal couples.** In the first mentioned class we have the copper-constantan[2] combination which, though it shows a high emf per degree (40 to 60 μv), tends to deteriorate at temperatures above 300°C. A couple of these metals can therefore be used only for the lower range of temperatures. In fact, this combination finds wide use at subzero temperatures.

In the same general class may be mentioned the iron-constantan couple. This combination has a straighter emf-temperature curve than

[1] See chapter on alternating currents.

[2] Constantan is an alloy consisting of 60 per cent copper and 40 per cent nickel.

the copper-constantan unit but has the defect that the iron may rust in a humid atmosphere. The iron-constantan couple may be used for observations up to about 800°C.

As a result of an effort to find a substitute for iron in thermocouples, Hoskins developed two alloys which have proven to be quite satisfactory, particularly for use at higher temperatures. One of these alloys is composed of 90 per cent nickel and 10 per cent chromium. The other metal consists of 98 per cent aluminum, 2 per cent nickel, and a trace of silicon and manganese. This combination is known under the trade name of the **chromel-alumel thermocouple.** The chromel-alumel pair may be used in measuring temperatures up to 1100°C continuously, and will function satisfactorily for short periods up to 1300°C.

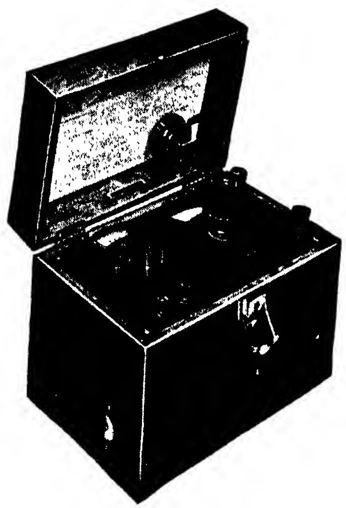


FIG. 132.—Portable potentiometer for use with thermocouple. (*Leeds & Northrup Co.*)

For still higher temperatures, and particularly where greater accuracy is essential, noble-metal thermocouples consisting of platinum and platinum-10 per cent rhodium are employed. These couples were introduced by Le Chatelier and are reliable up to 1500°C (2732°F).

There are two general methods in use by means of which a suitable thermocouple may be utilized in temperature measurements. In one, the couple is connected directly to a galvanometer or millivoltmeter, the indicating instrument being calibrated directly in degrees. Such an organization is referred to as a **thermoelectric pyrometer.**

In the plan just outlined the magnitude of the **thermoelectric current** is taken as an index of the temperature of the hot junction. In the second method, the **emf developed by the thermocouple** serves to indicate the temperature being observed. In order to measure conveniently and accurately the thermo-emf, recourse is had to the potentiometer, a device which we have already studied in Chap. IX, Sec. 79. The thermocouple and the associated potentiometer are known as a **potentiometer pyrometer.**

A simple and compact form of the potentiometer is used in this connection, the readings being made directly in degrees of temperature instead of volts. Figure 132 is an illustration of a well-known portable potentiometer type of temperature indicator, and Fig. 133 shows a panel

model. In both of the models the nul d-c potentiometer setting is made manually. In the panel type provision is made by means of the switches at the left for reading the temperature as indicated by several different thermojunctions which may be located at various points in a manufacturing plant.

There are also automatically operated potentiometer pyrometers actuated by thermojunctions. Figure 134 shows an indicating assembly of this type; and Fig. 135 is an illustration of a recording model. Frequently, in industrial applications, the thermojunction-generated emf not only serves to operate automatically the indicating and recording mechanism, but it also is made to function as a control agent. This is

FIG. 133.—Panel model of potentiometer type of temperature indicator. (*Leeds & Northrup Co.*)

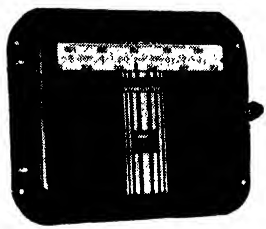


FIG. 134.—Indicating type of potentiometer pyrometer. (*Leeds & Northrup Co.*)

FIG. 135.—Recording model of potentiometer pyrometer. (*Leeds & Northrup Co.*)

brought about by incorporating the thermojunction in a balanced circuit. Any change in temperature in the region of the thermojunction results in a deflection of the associated galvanometer which in turn, by means of a relay and a valve-driven motor, serves to readjust the fuel supply, thus maintaining a constant temperature between narrow limits. Such a combined recording and control mechanism, though complicated, is very reliable, and is extensively used in industrial plants.

When used at high temperatures, thermocouples are enclosed in ceramic or metallic housings. Figure 136 shows a unit suitable for installation in the wall of a furnace.

Wherever conditions are such that the regular thermocouple assembly cannot be used, or wherever it is found desirable to observe the temperature of specific regions of a still or moving object, such as a hot metal sheet being rolled, use is made of a portable thermojunction assembly which is both convenient and reliable. Within a housing is assembled

a highly-sensitive thermopile and a suitable lens (or mirror) by which radiation from the hot object is brought to a focus on a nest of thermocouples. The emf thus developed is measured by means of a simple form of potentiometer that is calibrated to read in temperature units

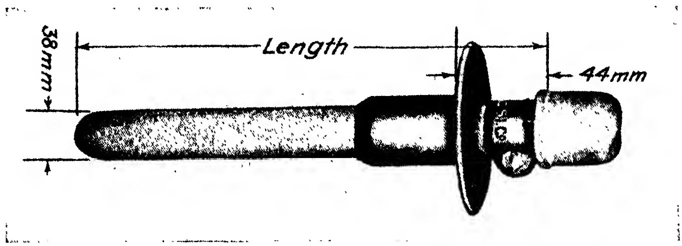


FIG. 136.—Thermocouple for use at high temperatures. (*Cambridge Instrument Co.*)

rather than in volts. A commercial unit (Fig. 137) known as a "Rayotube Pyrometer" finds wide application in industry. These units may be calibrated to cover several ranges between the limits of 225 to 3000°F.

108. Becquerel Effect. Bearing in mind that in this and the preceding chapter we are dealing with means whereby an emf may be developed, it will be in order to examine a phenomenon which involves the production of an emf as the result of the direct action of radiant energy. The generation of an emf in this case, however, apparently involves an electronic process that is different from that which obtains in connection with the Seebeck effect. In 1839, the French physicist, A. C. Becquerel, discovered that if two platinum or silver electrodes coated with silver chloride were immersed in dilute sulfuric acid, and one of the electrodes was illuminated, an emf was developed between the two electrodes. Such a phenomenon is now commonly referred to as a **photovoltaic effect.** Since Becquerel's original experiments, it has been found that the presence of an electrolyte is unnecessary. It is now known that certain metallic contacts will develop an emf when illuminated. Thus it becomes possible to transform radiant energy **directly** into electrical energy. In 1926 L. O. Grondahl discovered that the presence of light produced an effect on the action of a copper-oxide rectifier, which consists of a copper plate coated with a film of cuprous oxide. He suggested

FIG. 137.—Rayotube pyrometer. (*Leeds & Northrup Co.*)

that such a unit might have possibilities as a means whereby radiant energy could be transformed into electrical energy. Later (1930) Lange developed a light-sensitive unit that, when the surface is illuminated, develops an emf of sufficient magnitude to operate a sensitive relay.

The construction of the so-called copper oxide cell is indicated diagrammatically in Fig. 138. The Lange cell consists of a plate of copper, one side of which is covered with a thin layer of Cu_2O of the order of 10μ or less in thickness. On the cuprous oxide layer is deposited an exceedingly thin and semitransparent layer of copper. Electrical connection is made to the copper plate and the thin copper facing. When the copper film is illuminated, and connection is made between the two plates, an electronic current will flow from the copper plate to the copper film. The maximum open-circuit emf is of the order of 0.5 volt. If the external resistance R does not exceed a few ohms the relation between the intensity of the incident light and the magnitude of the current is linear. The unit can therefore be utilized as a photometric device and, in fact, is widely used for that purpose. Commercially the copper oxide unit is known as a **Photox cell.**

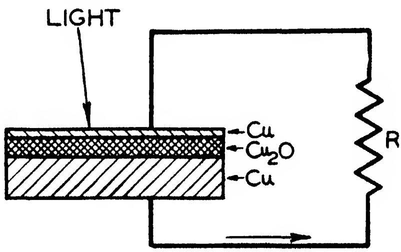


FIG. 138.—Copper oxide photoelectric cell.

Another example of this type of generator consists of a plate of iron over which is a thin coating of selenium mixed with a trace of silver. On the selenium is deposited a very thin metallic layer which serves as one terminal of the cell. This selenium-on-iron form of photoelectric cell functions in the same general manner as the copper oxide unit. Commercially it is referred to as the **Photronic cell.** The spectral response of the two types of cells just described is shown in Fig. 139. It is to be noted that the response of the copper oxide cell corresponds quite closely to that of the human eye.

The manner in which an emf is developed in a photovoltaic cell is still a matter of debate. The characteristics of the response curve (Fig. 139) would appear to preclude the possibility of the effect being a thermoelectric one. There are those who contend that we are here dealing with a case of photoelectric emission (Sec. 136); while others hold quite a different view. In any event, it would appear that the seat of the emf developed is the interface between the copper and its oxide in the first case, and the interface between the iron and the selenium in the second. Further research will undoubtedly lead to a better understanding of this important and widely utilized phenomenon.

109. Pyroelectric Effect. In addition to the Seebeck and the Becquerel effects there is at least one other method by which an emf may be developed directly from thermal energy. When certain crystals are heated they manifest an electrical charge, and this phenomenon is referred to as the **pyroelectric effect.** Among the crystals which exhibit this property may be mentioned tourmaline, quartz, fluor, and boracite. If a crystal of tourmaline, for instance, is heated, the ends will, during the process of heating, manifest opposite charges, and when cooling takes place the signs of the charges will be reversed. The initial tem-

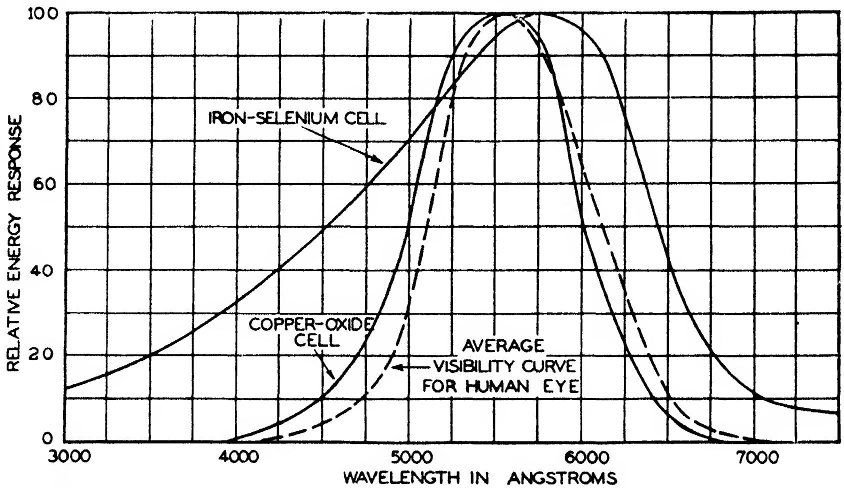


FIG. 139.—Response curves of photoelectric cells compared with that of the eye.

perature of the crystal does not affect the signs of the charges. If the crystal is broken into fragments, or even reduced to a powdered form, each part will exhibit the same characteristics.

110. Piezoelectric Effect. While the pyroelectric effect in crystals has, as yet, found no application, there is a procedure by which certain crystals may be made to develop quite an appreciable emf. In 1880, F. and P. Curie discovered a phenomenon[1] that, in recent years, has assumed great importance. Certain doubly-refracting crystals, notably quartz, tourmaline, and crystals of Rochelle salt,[2] exhibit electric charges on certain parts of the crystal when the crystal as a whole is subjected to mechanical pressure. This phenomenon is known as the **piezoelectric effect.** It was found by the Curies that the magnitude of the charge

[1] This subject is sometimes discussed in connection with the study of dielectrics. It is considered at this point in our discussion because we are dealing with methods of developing an emf.

[2] Rochelle salt is sodium potassium tartrate, $NaKC_4H_4O_6 \cdot 4H_2O$.

thus made manifest is proportional to the pressure, and may be represented by the relation $Q = KP$, where Q is the charge, P the total force applied to the crystal, and K a constant known as the piezoelectric constant. Thus if an alternating mechanical pressure is applied to such a crystal a corresponding alternating emf will be developed. Further, **the response is independent of the frequency between very wide limits.** A thermojunction when subjected to wide temperature differences develops an emf measured in millivolts. A crystal that exhibits the piezoelectric property to a marked degree will produce an emf of the order of a volt when subjected to a pressure of a few grams. This is particularly true of Rochelle salt crystals. Nicholson[1] and others have thoroughly investigated the piezoelectric effect in the case of this crystal, particularly in connection with the possible utilization of this phenomenon in the field of electroacoustics. A salt crystal may be easily arranged to show the piezoelectric effect by compressing the crystal between two metal plates applied to two opposite faces and arranging a metallic band around the middle of the crystal. If the two pressure plates are electrically connected to form one electrode and the central band serves as the other connection and the system thus formed is connected to a potential-measuring device, it will be found that very slight variations in pressure will develop a variable emf of considerable magnitude. Indeed, under suitable conditions, the pressure due to sound waves may be made to produce a terminal emf of several millivolts. The piezoelectric effect is utilized in the design of phonograph reproducers and sound pickups (so called "crystal microphones"). A Rochelle salt crystal when mechanically articulated with a properly mounted phonograph needle will be subjected to a variable pressure as the needle follows the groove of the recording. Thus the mechanical recording will be transformed into a varying potential having an average value of the order of 1 volt. A Rochelle salt crystal having dimensions of the order of 1.8 cm long, 1.2 cm wide, and 0.76 mm thick is used for such a purpose.

In common with all mechanical systems, such a crystalline structure has a natural period of mechanical oscillation. The dimensions of the crystal used for acoustical purposes are so chosen that its natural frequency lies above the useful frequency range. It is possible to design a phonograph pickup which will have a reasonably flat response between 40 and 8,000 cycles/sec. Temperature variations modify slightly the emf developed, though not seriously. Piezoelectric phonograph pickups are widely and successfully used.

When used as sound-pickup units, the Rochelle salt crystal is usually slightly smaller in size than those used in a phonograph assembly. A

[1] *Proc. AIEE*, p. 1,467 (1919).

representative section would be of the order of 1 cm^2 and 0.5 mm thick. Since the pressure exerted by incident sound waves is relatively small, several crystal elements are often connected in series, or in a series-parallel grouping. By allowing the sound waves to impinge directly on the crystal elements a very flat response between 60 and 6,000 cycles/second is secured, the response rising slightly from that point as the frequency approaches 10,000. However, special microphones of this type have been developed which give a satisfactory response up to the limit of audibility. Sound pickups are made in which the sound waves are received on a diaphragm which in turn serves to convey the pressure to a crystal element. While this type of crystal microphone gives a higher emf output than the type just mentioned, its frequency response is not as satisfactory.

Fig. 140.—Photograph of a quartz crystal. Note the geometrical form.

Rochelle salt crystals are utilized in electroacoustical pickup devices because of the relatively high emf developed for a given pressure. In the case of quartz the piezoelectric constant is only one-thousandth that of Rochelle salt; but quartz is much more stable mechanically, and is accordingly used where such a crystal element would be subjected to severe mechanical treatment. Its lower emf output does not constitute a serious limitation because dependable high-gain amplifiers (Sec. 236) are readily available.

In addition to the applications above cited, use has been made of the piezoelectric effect in the design of pressure and surface gauges, and other similar devices.

In their study of piezoelectric phenomena the Curies discovered that the effect which we have just considered is reversible, *i.e.*, if a crystal that shows the **direct piezoelectric effect** is subjected to a potential difference between its faces, its crystalline structure will undergo a change in thickness and other, though lesser, dimensional changes will occur. The latter behavior is sometimes referred to as a **converse piezoelectric effect.** If the electric field to which the crystalline plate is subjected is of an alternating character, the plate will undergo a corresponding mechanical deformation. Since this important property is widely util-

ized in communication engineering, we will examine this aspect of the piezoelectric effect somewhat in detail.

Quartz, because of its mechanical properties, is almost universally employed when use is made of this reverse property. Natural quartz (SiO_2) crystals have the appearance indicated in the photograph shown in Fig. 140. In Fig. 141 it is shown that the body cross section is hex-

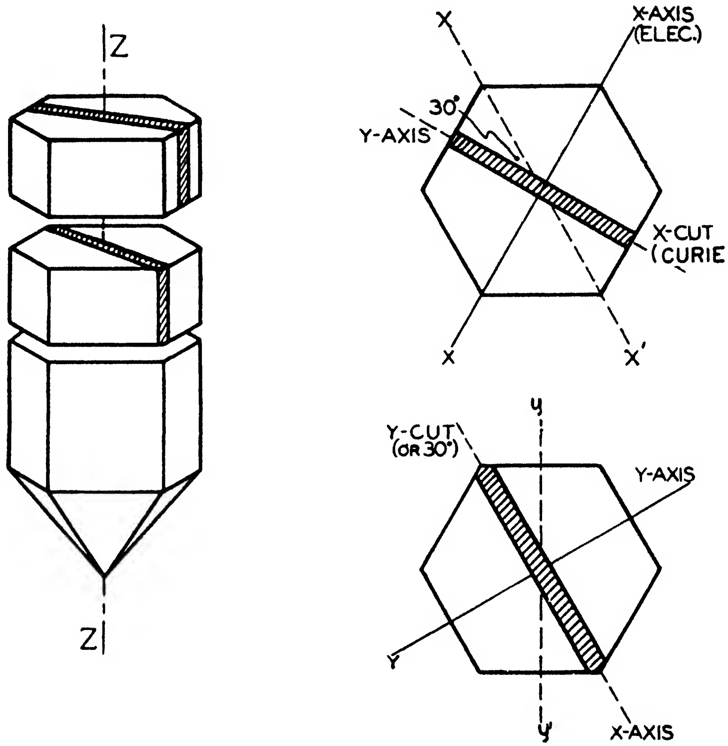


FIG. 141.—Showing the geometrical relation of piezoelectric sections to the several axes of a quartz crystal.

agonal in shape and that there are three sets of mutually perpendicular axes. The electromechanical properties of a section cut from such a crystal depend upon the spatial relation which the particular section bears to these three axes. The line Z joining the apices of the crystal is the optic axis. A line cutting two opposite faces of the hexagon, and normal to the optical axis, is referred to as a Y-axis, or a mechanical axis; and a line bisecting any two opposite angles, and perpendicular to the optic axis, is considered to be an X-axis, or an electrical axis. Obviously there are three X-axes and three Y-axes. These relations are indicated in the illustration appearing as Fig. 141.

If a flat section having its faces normal to an X-axis (Fig. 141) is cut from a quartz crystal, we have what is known as an X, or **Curie cut.** If a mechanical stress is caused to exist along the Y-axis of such a section, electrical charges of opposite polarity will appear on the flat sides of the section. If the stress is changed from a compression to a tension, for instance, the sign of the charges appearing will be reversed. And, conversely, if a potential difference is established between the two faces of the section, a mechanical stress will be manifest along the X-axis. Likewise, if a section is to be cut with its faces normal to the Y-axis it is referred to as a Y or **30° cut** (Fig. 141). The electromechanical behavior of such a section is, in general, similar to that which obtains in the case of the X cut. A section whose faces are normal to the optic axis does not show any piezoelectrical effect. (Compare this behavior with the corresponding optical characteristics of such a crystalline structure.)

If conditions are so arranged that an alternating potential difference is established between the two flat faces of either an X or a Y section, as indicated in Fig. 142, the physical structure of the crystal will undergo periodic deformation; and these structural oscillations will be in synchronism with the impressed potential. If the frequency of the applied alternating potential approximates the natural mechanical frequency of the crystal, the amplitude of the structural oscillations will be relatively great. Indeed, if resonance obtains and the impressed potential difference is great enough, the crystal may break into a number of pieces.

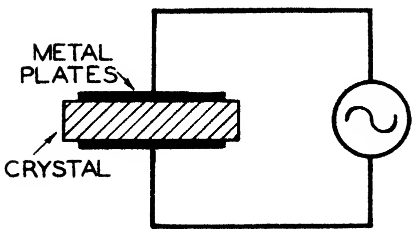


FIG. 142.—Oscillating piezoelectric plate.

While both the X- and Y-cut sections may be caused to oscillate, as outlined above, their detailed mechanical behavior differs somewhat. For instance, Y-cut plates will usually oscillate more readily than will the X-cut sections, particularly at the lower frequencies. The resonating properties of quartz crystals were discovered by Dr. W. G. Cady.

Temperature has an effect on the natural frequency of piezoelectric crystals. X-cut plates show a negative temperature coefficient, the frequency **decreasing** with rise of temperature. The frequency of such a section will change something like 15 to 25 cycles in a million for each centigrade change in temperature. In the case of a Y-cut plate the temperature coefficient is positive, the frequency change being from about 20 to 100 cycles per megacycle per degree centigrade. Morrison[1] has

[1] MORRISON, W. A., A High Precision Standard of Frequency, *Proc. IRE*, July, 1929.

found that if one cuts an annular ring from a Y-cut plate the two temperature coefficients tend to cancel out approximately. It has also been determined that[1] if a plate is cut from a parent crystal at an angle of 35° to the ordinary Y-cut position, its frequency is almost completely independent of temperature change. Such a section is known as an AT cut, its relation to the other cuts being shown in Fig. 143.

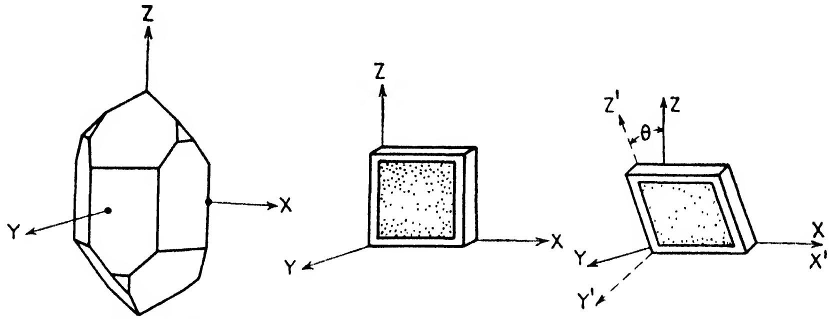


FIG. 143.—Showing the orientation of an AT-cut plate to the X, Y, and Z axis of a quartz crystal. (*Courtesy Bell Laboratories Record.*)

Young's modulus in crystalline quartz varies with direction, and in a complex manner[2]. This would lead one to expect that a given section of such a crystal might manifest a number of frequencies simultaneously; and such is found to be the case. The frequencies which do obtain in any given instance depend upon the size of the plate, the orientation of the section with respect to the several axes, and the manner of excitation. The modes of vibrations may be longitudinal, transverse, or torsional. Elastic coupling between the several types of vibrations exists; hence combinations of these modes sometimes occur. In the case of quartz plates, one of the higher natural frequencies is commonly utilized in practice. In such cases the thickness of the section is the governing factor, the frequency being proportional to that dimension. An X-cut plate having a thickness of 1 mm will have a frequency of approximately 3 megacycles, and a Y-cut section of the same thickness would show a frequency of about 2 megacycles. A working relation which is often used in the cutting and grinding of such crystalline sections is, for the X cut,

$$f_X = \frac{112.6}{t}, \tag{137}$$

[1] WILLARD, G. W., Elastic Vibrations of Quartz, *Bell Lab. Record*, April, 1936.

[2] WILLARD, *loc. cit.*, and F. R. LACK, Observations on Modes of Vibration and Temperature Coefficients of Quartz Crystal Plates, *Proc. IRE*, July, 1929.

where f_X is the frequency in megacycles and t the thickness in thousandths of an inch. For a Y-cut plate the relation is

$$f_Y = \frac{77.0}{t}. \tag{138}$$

The frequency of an AT cut is given by

$$f_{AT} = \frac{662}{t}. \tag{139}$$

The manner in which plates cut from natural quartz crystals are excited, and the uses to which oscillating units are put, will be discussed in a later chapter. For the present it may be said that the frequency of all radio transmitters is governed by means of such a crystal, supersonic waves are generated by their use, they enter into the construction of highly accurate timepieces, and they are incorporated in the design of electrical filters. Thus it will be seen that wide practical use is made of the piezoelectric effect—a phenomenon which for many years was but an interesting scientific curiosity. Today the supply of natural quartz crystals scarcely meets the demand. Possibly some enterprising chemist will eventually be able to produce quartz crystals from fused quartz.

CHAPTER XV

MAGNETIC EFFECTS OF THE ELECTRIC CURRENT

111. The Oersted Effect. Having made a study of the thermal and chemical effects of the electric current we shall next consider the magnetic effects which result from the movement of electrons. Thus we enter the domain of electrodynamics.

From time to time in the history of science investigations have been undertaken which have led to epoch-making discoveries. Such an experiment was performed by Professor Hans Christian Oersted in 1819, and announced by him in the following year. By holding a wire, through which a current was flowing, parallel to a magnetic needle, Oersted found that the needle was deflected, thus clearly establishing the enormously important fact that an electric current gives rise to a magnetic field. His experiment also showed that the magnetic field at any given point, due to the current in the wire, is perpendicular to the direction of the current. It may truly be said that the science of what we shall call electromagnetism had its beginning in Professor Oersted's classical experiment. Oersted's discovery united the fields of magnetism and electricity, and thus opened a field having far-reaching scientific and practical possibilities.

Oersted's announcement excited widespread interest, and resulted in energetic experimentation on the part of many other investigators. Within a few months important advances were made in this new field. In order to increase the magnetic effect due to a given current, M. Andre Ampere conceived the plan of forming the conducting wire into a spiral or helix. He also discovered that two conductors, each of which is carrying a current, will react magnetically upon one another without the presence of a magnetic material; and further that such parallel currents, when flowing in the same direction, attract one another, and when in opposite direction exhibit repulsion. We shall see that these observations of Ampere have been important factors in the development of the science of electrodynamics.

However, before proceeding to consider the quantitative relations that obtain between the magnitude of the current and the field produced, it will be well to observe that an electromagnetic effect is produced **only when the electrons** that constitute the current **are in motion.** A single

electron in motion gives rise to a magnetic field, but a thousand coulombs **at rest** would produce absolutely no magnetic effect.

Rowland, in 1876, demonstrated that a moving charge does produce a magnetic field. He arranged to rotate rapidly an insulating disk on the surface of which had been placed electric charges. He found that a small magnet suspended near the rotating disk was deflected, as it would have been if an electric current had existed in the region occupied by the moving charges.

It should be observed that the magnetic effects due to electrons in motion are entirely independent of the nature of the conductor. For instance, an isolated electron, whether being carried on a Rowland rotating disk, or moving along a copper wire, or through the highest possible vacuum, will give rise to a perfectly definite magnetic field. The magnitude of the magnetic effect produced in either case will depend upon certain definite factors, **but the character of the conductor is not one of them.**

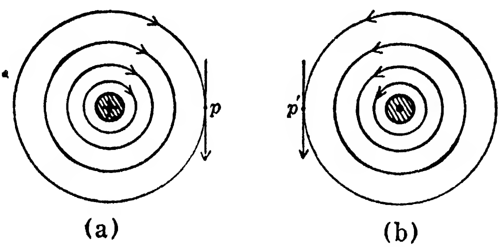


Fig. 144.—Relation of magnetic flux to the direction of the current.

One other fact, noted soon after Oersted's discovery, should also be mentioned, viz., that the lines representing the magnetic field about a conductor form concentric circles about the direction of the current as an axis. For example, if the stream of electrons constituting the current is moving in a direction normal to this page and toward the reader, as shown in Fig. 144*a*, the lines of force will be as shown by the concentric circles and the general direction of the field will be clockwise as shown by the arrows. If the current is away from the reader, the field will be as shown in Fig. 144*b*. If the **left** hand is closed, and the thumb points in the direction of the electronic current, the fingers will point in the direction of the magnetic flux. (If the conventional current is being considered the **right** hand rule holds.) At any particular point p the direction of the magnetic field will be as indicated by the straight arrows.

112. Laplace's Rule. Within two months after the announcement of Oersted's discovery two French physicists, J. B. Biot and F. Savart, reported to the French Academy that they had discovered the relation

that exists between the magnetic field intensity at a point, produced by a current flowing in a long straight wire, and the distance of the point from the conductor. Biot and Savart found that the field intensity varied directly as the current strength and inversely as the distance from the conductor carrying the current. Laplace, an eminent French analyst and astronomer, in discussing the law established by Biot and Savart for the special case examined by them, showed that a generalization might be formulated which would take the form

$$dH = \frac{I\,dl \sin\phi}{r^2}, \tag{140}$$

where dH (Fig. 145) represents the element of field strength at p due to the current element dl, r the distance of the point p from the middle of the element, and ϕ the angle which the element dl makes with the line r. The direction of the field dH is normal to the plane of the figure at p. This generalization has come to be known as Laplace's rule; it will serve as a basis for important deductions to follow.

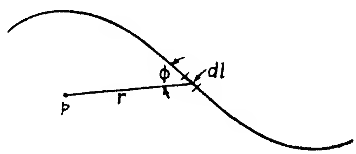


FIG. 145.—Laplace's rule.

113. Field Due to a Linear Current. The results enunciated by Biot and Savart may be readily deduced by employing Laplace's general relation [Eq. (140)]. Suppose we have a conductor of infinite length through which a current I is flowing; let us find an expression for the field intensity at any point P distant x centimeters from the nearest point S on the conductor, as set forth in Fig. 146. In making use of Laplace's equation it will be more convenient, in our case, to deal with angle θ than with angle ϕ. Since $\sin\phi = \cos\theta$ we may write Laplace's equation in the form

$$dH = \frac{I\,dl \cos\theta}{r^2} \qquad \text{oersteds.}$$

We can eliminate dl, and also express r in terms of x. From the similar triangles QMN and MPS we may write

$$\frac{dl}{MN} = \frac{r}{x}.$$

But

$$MN = r\,d\theta.$$

Hence

$$\frac{dl}{r\,d\theta} = \frac{r}{x},$$

or

$$dl = \frac{r^2\, d\theta}{x}.$$

Substituting in our form of Laplace's equation, and simplifying we have

$$dH = \frac{I\, d\theta \cos\theta}{x},$$

as the contribution of the current element dl to the field intensity at P. The **total** field at P will be given by

$$\int dH = \frac{I}{x}\int_{\theta=-\theta_2}^{\theta=\theta_1} \cos\theta\, d\theta = \frac{I}{x}\Big[\sin\theta\Big]_{-\theta_2}^{\theta_1}$$

$$H = \frac{I}{x}[\sin\theta_1 - \sin(-\theta_2)]$$

$$= \frac{I}{x}(\sin\theta_1 + \sin\theta_2).$$

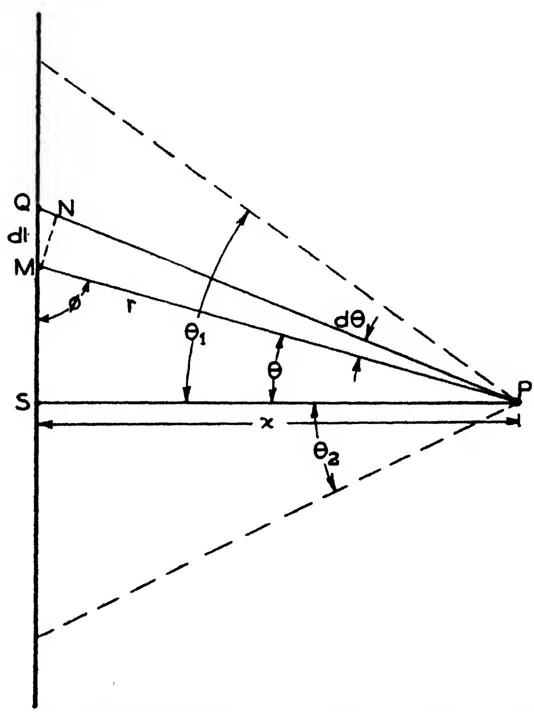


FIG. 146.—Magnetic field at a point due to a linear current.

For an infinitely long wire

$$\theta_1 = \theta_2 = \frac{\pi}{2},$$

and hence

$$\sin \theta_1 = \sin \theta_2 = 1.$$

Therefore the total field at P due to the current in an **infinitely long** straight conductor will be given by

$$H = \frac{2I}{x}, \tag{141}$$

where H is the field strength in oersteds; I the current in abamperes;[1] and x the distance in centimeters from the point P to the current path, which is in conformity with the observations of Biot and Savart. This means that if a unit test pole is placed at a distance x from an infinitely long wire carrying a current it will experience **a mechanical force in dynes** given by $2I/x$. The direction of the field at P is normal to the plane of the figure. Obviously this expression will also give the approximate field strength at a point **very near** to a wire of **finite** length.

114. Field at any Point on the Axis of a Circular Loop. Assume a circular loop carrying a current I, as shown in Fig. 147. Our problem is to find the field intensity at any point P on the axis of the loop. Laplace's relation [Eq. (140)] may be used to advantage in this case also.

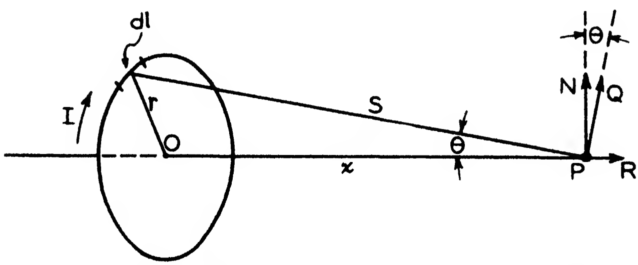


FIG. 147.—Magnetic field on the axis of a circular loop.

Consider first the contribution made by an element of the circuit dl to the field at P which is at a distance x from the center of the loop. Since the line S is normal to dl, Laplace's equation becomes, in this case,

$$dH = \frac{I\,dl}{S^2}.$$

dH will have the direction of PQ. We are interested, however, in the field parallel to the axis OP. The field in the direction PQ may be

[1] This unit of current is defined in Sec. 115.

resolved into components along PN and PR. If the entire loop is considered, the elemental components normal to OP will cancel out in pairs, leaving the vector sum of the dH components (along PR) as the resultant field.

The component parallel to OP due to dl will be given by

$$dH = \frac{I\,dl}{S^2}\sin\theta$$
$$= I\frac{dl}{S^2} \times \frac{r}{S}.$$

The total field will be

$$\int dH = \frac{Ir}{S^3}\int_0^{2\pi r} dl,$$

which leads to

$$H = \frac{2\pi I r^2}{S^3}.$$

We may express S in terms of x and r, and get

$$H = \frac{2\pi r^2 I}{(x^2 + r^2)^{3/2}} \quad \text{oersteds.} \tag{142}$$

If P is at the center of the loop, Eq. (142) becomes

$$H = \frac{2\pi I}{r}, \tag{143}$$

which gives the field intensity in oersteds at the center of a single loop the radius of which is r.

Further, if the loop consists of several turns instead of one, the turns being connected in series and occupying a space small compared with the radius, the field at the center will be given to a close approximation by

$$H = \frac{2\pi I n}{r} \quad \text{oersteds,} \tag{144}$$

where n is the number of turns making up the coil.

115. Electromagnetic Unit of Current. In applying Eqs. (141) to (144) to actual cases, we are at once confronted with the question of units. These equations are different expressions for field intensity **in terms of force action on unit magnetic pole.** We have already seen (Sec. 20) that, when dealing with numerical values, the cgs unit of magnetic field intensity is expressed in dynes per unit pole. In the equations referred to, distances will of course be expressed in centimeters. It

therefore becomes necessary to set up a definition of the cgs unit of current.

Following the usual custom in such cases we may make all the independent variables in the defining equation unity. In deriving Eq. (143) we integrated over the entire loop. If, instead of doing this, we had made the radius of the loop unity, and had considered only **unit length of the conductor,** the field intensity H would be **numerically** equal to the current I. It may therefore be said that **the cgs unit of current is defined as a current of such magnitude that, when flowing in a conductor in the form of an arc whose radius is one centimeter and whose length is one centimeter, it will exert a force of one dyne on unit magnetic pole placed at the center of the arc.** Since this unit is based on unit magnetic pole it is known as the cgs electromagnetic (em) unit of current, or the **abampere.**

In Sec. 19 the practical unit of current, the ampere, was defined in terms of electrostatic units. The ampere may also be defined in terms of the em unit of current (the abampere). For reasons which we will examine later, it has been agreed to say that the ampere is equivalent to one-tenth of the em unit of current, *i.e.*, to one-tenth of an abampere. It therefore follows that, if I in Eqs. (141) to (144) is to be expressed in amperes, the factor 10 should appear in the denominator of each of those equations.

Problem. A current of 2 amp is flowing in a conductor which forms a circular loop the diameter of which is 14 cm. What is the magnetic field strength at a point on the axis of the loop 10 cm from the loop's center?

Solution. $$H = \frac{2\pi r^2 I}{10(x^2 + r^2)^{3/2}} = \frac{2\pi 7^2 \times 2}{10(10^2 + 7^2)^{3/2}} = 0.0662 \text{ oersteds.}$$

Problem. A closely wound coil consisting of 100 turns has a mean diameter of 10 cm. If a current of 2 amp flows through the wire, what is the field strength at the center of the coil?

Solution. $$H = \frac{2\pi I n}{10r} = \frac{2\pi \times 2 \times 100}{10 \times 5} = 8\pi \text{ oersteds.}$$

116. Helmholtz Coils. In certain important cases, several of which occur in the study and applications of electronics, it is necessary to be able to produce a magnetic field which is **uniform** over a distance of at least a few centimeters. In order to develop such a field, recourse is had to an electromagnetic system consisting of two coils arranged as originally used by Helmholtz in connection with a type of galvanometer that is now more or less obsolete.

If a region can be found, in the field due to a single coil, where the **rate**

of change in field strength is **constant,** it would then be possible so to combine the fields due to **two** similar coils that a uniform field would obtain, at least in a restricted region.

We may learn whether such a region of constant rate of change exists by examining the relation embodied in Eq. (142). For a coil having n turns this expression becomes

$$H = \frac{2\pi r^2 In}{(x^2 + r^2)^{3/2}}.$$

If there is a point where the rate of change of field strength with respect to the axial distance from the center of the coil is constant, it may be found by equating the second differential of the x term in the above expression to zero. Thus

$$\frac{d^2}{dx^2}\left[\frac{1}{(x^2 + r^2)^{3/2}}\right] = 0.$$

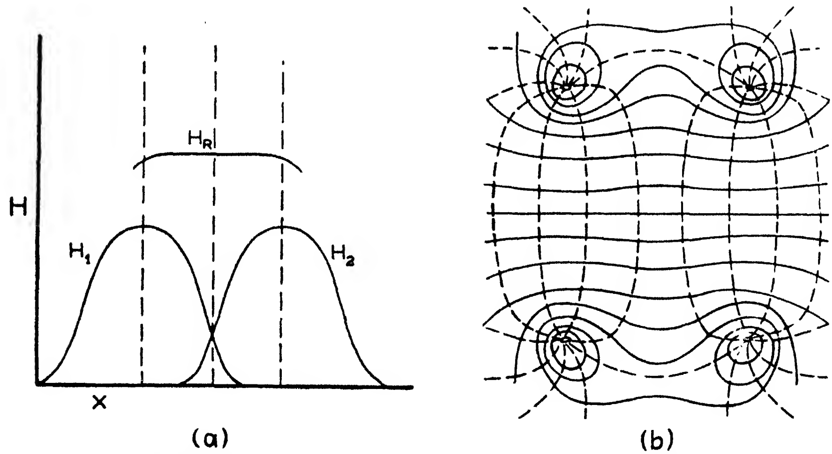


FIG. 148.—Magnetic field in the region between a pair of Helmholtz coils.

This procedure leads to $x = r/2$; which means that, at a point equal to half the radius of the coil, the rate at which the field strength H changes is constant. If, then, two similar coils, connected in series, are arranged coaxially at a distance from one another equal to their radius, the decrease in the field of one of the coils will be compensated by the increase of the field due to the other, at least over a limited region. This is shown graphically in Fig. 148*a*. The curves H_1 and H_2 represent the field intensities as a function of x, while H_R shows the resultant field due to the two coils. The horizontal part of the H_R curve indicates the region in which the field strength is approximately constant. Figure 148*b* roughly

illustrates the field pattern in the region between the two Helmholtz coils.

At a point on the common axis midway between the two coils the magnetic field strength will be given by the expression

$$H = 2\frac{2\pi n r^2 I}{\left(\frac{r^2}{4} + r^2\right)^{3/2}} = \frac{32\pi n I}{5^{3/2} r} = \frac{2.86\pi I}{r} \quad \text{oersteds,} \tag{145}$$

where I is in abamperes and r is the radius of the coils in centimeters.

117. Field Intensity within a Helix. We may extend the results of Sec. 114 to include the case of the solenoid. An helical winding is a form of circuit frequently met with in practical electrical equipment; its magnetic properties are therefore of peculiar interest. Let us suppose that we have a solenoid the wire of which is wound in a single layer having n

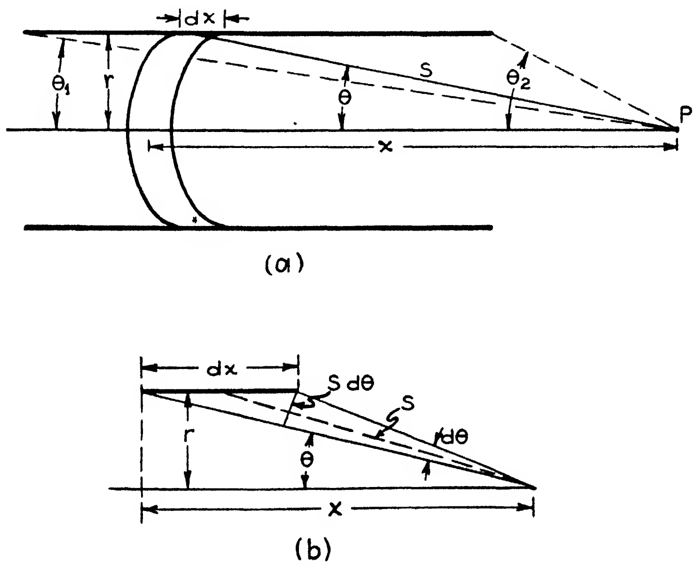


Fig. 149.—Field intensity within a helix.

turns **per unit length.** Our problem consists in finding an expression for the field intensity at any point on the axis as shown in Fig. 149a. Consider a very small section of the winding dx. This section will contain $n\,dx$ turns. The field at P, due to the element dx, is (Sec. 114)

$$dH = \frac{2\pi I r^2 n\,dx}{S^3}. \tag{i}$$

It will facilitate the evaluation of this relation if we express dx in terms of θ. Referring to Fig. 149b, we may write

$$\frac{S\,d\theta}{dx} = \sin\theta,$$

or

$$dx = \frac{S\,d\theta}{\sin\theta}. \tag{ii}$$

Further,

$$\sin\theta = \frac{r}{s}. \tag{iii}$$

Substituting the values from (ii) and (iii) in (i), we get

$$dH = 2\pi In \sin\theta\,d\theta.$$

The field at P due to the complete helix will be

$$\int dH = 2\pi nI \int_{\theta=\theta_2}^{\theta=\theta_1} \sin\theta\,d\theta,$$

where θ_1 and θ_2 are values of θ at the ends of the solenoid. This gives

$$H = 2\pi nI \Big[\cos\theta\Big]_{\theta_2}^{\theta_1}$$
$$= 2\pi nI(\cos\theta_1 - \cos\theta_2). \qquad \text{oersteds} \tag{146}$$

If the point P is moved to the nearest end of the helix, $\theta_2 = 90°$ and hence $\cos\theta_2 = 0$; thus giving for Eq. (146)

$$H = 2\pi nI\cos\theta_1.$$

Expressing the angle θ_1 in terms of the length of the solenoid l and its radius r, the above equation becomes

$$H = 2\pi nI\frac{l}{\sqrt{r^2 + l^2}} \qquad \text{oersteds,} \tag{147}$$

which gives the magnetic field intensity at the end of the solenoid, when I is in abamperes.

If the length of the helix is great in comparison with the radius, Eq. (147) becomes

$$H = 2\pi nI. \qquad \text{oersteds} \tag{148}$$

If P is at the center of the solenoid, the angles θ_1 and θ_2 are equal and, for a very long slender coil, approach zero as a limit. Hence

$$\cos\theta_1 - \cos\theta_2$$

becomes sensibly equal to 2, and Eq. (146) reduces to

$$H = 4\pi nI \qquad \text{oersteds,} \tag{149}$$

which is the field intensity at the center of a long slender helix. A comparison of Eqs. (148) and (149) shows that the field strength at the ends of the winding is only half what it is at the center.

If the current I is in amperes, Eqs. (148) and (149) become

$$H = \frac{2\pi nI}{10} = 0.628nI \qquad \text{oersteds,} \tag{150}$$

and

$$H = \frac{4\pi nI}{10} = 1.256nI \qquad \text{oersteds,} \tag{151}$$

respectively.

In the case of a **torodial winding** (Fig. 150) the field is uniform throughout the interior of the coil, and hence Eq. (151) applies at any central point within the winding. If the solenoid is made up of a multiple-layer winding, as is commonly the case in practice, and if the thickness of the winding is small compared with the radius of the coil, Eq. (151) will give the field intensity to a first approximation. In such a case the mean radius is used in the computation.

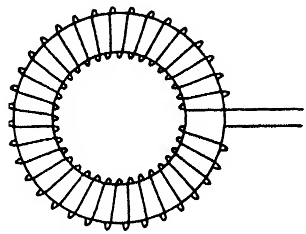


FIG. 150.—Torodial winding.

Problem. A solenoid consisting of 500 turns is 20 cm in length. What will be the field intensity near the center of the coil if it carries a current of 5 amp?

Solution. Since the number of turns is 500 and the length 20 cm, the number of turns per centimeter will be 25. Employing Eq. (151) we have

$$H = 1.256nI = 1.256 \times 25 \times 5 = 157.07 \text{ oersteds.}$$

118. Mechanical Force on a Current in a Magnetic Field. We have seen that a current gives rise to a definite magnetic field and that the intensity of that field may be easily calculated (Sec. 113). The magnitude of the field was defined in terms of the **mechanical force experienced by unit test pole at the point in question.** If the current exerts a force on a magnetic pole, **the pole also exerts a force on the current,** as we would expect from Newton's third law. Since this mutual force action between currents and magnetic fields lies at the basis of the operation of many electrical meters, motors, and other similar electrical equipment, it is important that we examine the basic relations that obtain between the several factors involved. Fortunately the relations are simple in form and direct in their application.

Let us first derive an expression for the mechanical force experienced

by a conductor of unit length, carrying a current, and located in a magnetic field due to some agency outside the conductor itself. Referring again to a simple loop (Fig. 151), the field at the center, as given by Eq. (143), is $H = 2\pi I/r$. If a unit magnetic pole is placed at the center of the loop it will experience a mechanical force of $2\pi I/r$ dynes. If a pole whose strength is m be similarly placed, the force will be

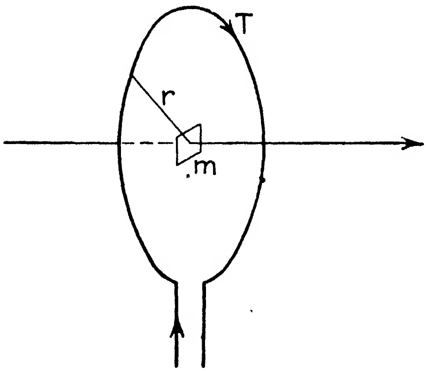


Fig. 151.—Mechanical force on a current in a magnetic field.

$$F = \frac{2\pi I m}{r} \qquad \text{dynes.}$$

As implied above, such a magnetic pole will likewise exert an equal force on the current and hence on the conductor forming the loop. This force is exerted by the magnet **on the loop as a whole.** The force experienced by **unit length** of the loop would be

$$\frac{2\pi I m}{r} \times \frac{1}{2\pi r} = \frac{Im}{r^2},$$

because the total length of the conductor is $2\pi r$. It follows from the fundamental law of magnetism (Sec. 19) that the field intensity H at the loop, **due to the magnet at the center of the loop,** will be given by m/r^2. Hence the **mechanical force experienced by unit length**[1] **of the conductor** may be expressed in the form

$$F = IH \text{ dynes/cm length,} \tag{152}$$

where I is in abamperes. If we express I in amperes, and deal with a conductor whose length is l, the above relation takes the form

$$F = \frac{IlH}{10}, \tag{153}$$

where F is in dynes and l in centimeters. It is left for the student to show that, if F is expressed in pounds of force, l in inches, and B in lines per square inch, the above expression will take the form

[1] In the case of a loop, any section of the conductor, however small, will not be straight. As r increases, however, unit length will approach a straight line; and since Eq. (143) applies to a loop of any size, Eq. (152) will give force on unit length of a straight conductor.

$$F = \frac{BlI}{1.13 \times 10^7}. \tag{154}$$

This relation is useful in connection with the design and operation of d-c motors; and indeed it represents one of the basic laws in electrodynamics.

The results just derived may be extended to include a form of circuit frequently met with in practice, namely, a rectangular coil. Let the conditions be as illustrated in Fig. 152*a*. The side AD will experience a

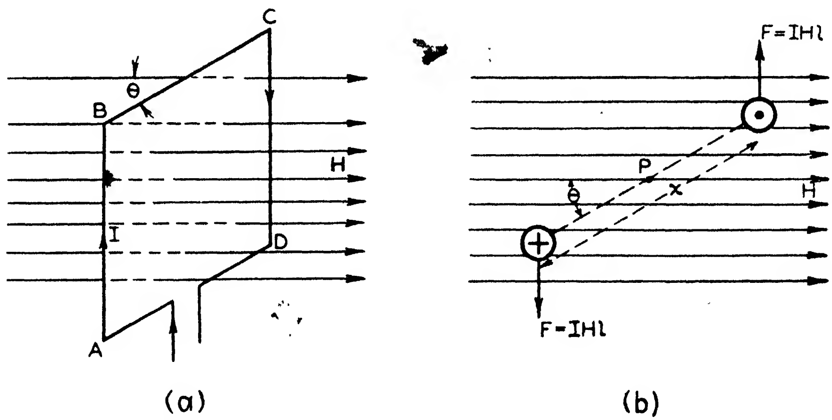


Fig. 152.—Couple due to a current in a rectangular loop located in a magnetic field.

force $IH(AD)$ tending to move it upward and the side BC will be acted upon by an equal force $IH(BC)$ directed downward; hence the resultant force on these two sides will be zero. The force action on the sides AB and CD will be as diagrammed in Fig. 152*b* when viewed from the side AD. If the length of AB and CD equals l, the force F operating on each side will be IHl dynes. It is evident that we have a couple acting on the coil and tending to produce rotation about an axis through P. The value of this couple will be given by

$$L = IHl(x \cos \theta), \tag{155}$$

where L represents the couple in dyne-cm; θ the angle that the coil makes with the direction of the field; and x the width of the coil, in this case AD or BC. It will be noted that the product xl gives the area of the coil; hence we may write

$$L = IHA \cos \theta, \tag{156}$$

where A is the area of the coil. If the coil consists of a number of turns in series, each loop would contribute its part to the total torque action; hence a general expression for the couple on a coil in a magnetic field will take the form

$$L = IHAN \cos \theta, \tag{157}$$

where N is the number of turns in the coil, A the average effective area of the individual turns, and I the current in abamperes. When $\theta = 0$, L will have its maximum value. This condition obtains when the plane of the coil is parallel to the direction of the field. It is therefore evident that the coil will tend to turn so that the area within the coil will include the greatest number of lines of force. If I is in amperes the last expression becomes

$$L = \frac{IHN \cos \theta}{10} \quad \text{dyne-cm.} \tag{158}$$

119. Galvanometers. The relations derived in the last section find application in several important connections, the most notable of which are the moving-coil galvanometer and the d-c motors. For the present we shall consider the theory of the D'Arsonval type of galvanometer and reserve for a later chapter a discussion of motors. In 1882, D'Arsonval[1] developed a current-indicating instrument which is now almost universally used where direct currents are involved. In this form of current-indicating device, a rectangular coil of many turns is suspended in the field of a permanent magnet by means of a delicate gold, phosphor bronze, or steel strip (Fig. 153). The suspension strip connects electrically with one terminal of the coil winding and serves as one connection to the source of current. A coiled metallic strip connects the other terminal of the coil to the source. A mirror or pointer attached to the coil serves to indicate any rotation of the moving system. The coil is free to move about a fixed soft-iron core, the latter serving to increase the field strength in the region of the winding.

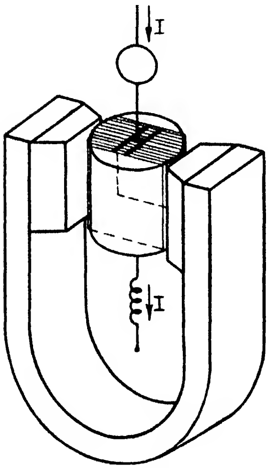


FIG. 153.—D'Arsonval type of galvanometer.

When current is passing through the winding, the coil will turn about its axis in response to the electromagnetic couple which will be developed; and the magnitude of the deflecting moment will be given by Eq. (158). This electromagnetic couple will be opposed by a restoring moment due to the twisting of the suspension. The coil will come to rest when the

[1] The suspended-coil type of instrument was first used by Sir William Thomson in 1870. However, this form of galvanometer did not come into extended use until D'Arsonval introduced the design of moving-coil instrument that bears his name.

deflecting moment equals the restoring moment. From the principles of mechanics we know that the moment due to the torsion of the suspension is proportional to the angular displacement of the end attached to the coil. This mechanical moment will be given by $T_o\theta$, where T_o is the moment of torsion of the suspension and θ the angular displacement of the coil in radians. The value of T_o is any given case will obviously depend on the physical characteristics of the suspension. We may therefore write

$$T_o\theta = \frac{IHAN}{10}\cos\theta$$

$$\theta = \frac{IHAN}{10T_o}\cos\theta. \tag{159}$$

The last equation shows that the magnitude of the deflection produced by a given current is a function of the field strength, the number of turns constituting the winding, and the mechanical characteristics of the suspension. However, because of the cosine term the relation is not a linear one. In many cases it is convenient to have available a current-indicating instrument in which a linear relation does obtain between the rotation of the moving system and the magnitude of the current passing through the coil. In order to accomplish this the magnetic system of the galvanometer is designed as shown diagrammatically in Fig. 154. As a result of the presence of the soft-iron core centrally located with respect to the curved pole pieces, the magnetic flux in the air gap is radial. Under these circumstances, the plane of the coil will always be, within relatively wide limits, parallel to the direction of the flux. This means that $\cos\theta = 1$, and our working equation accordingly becomes

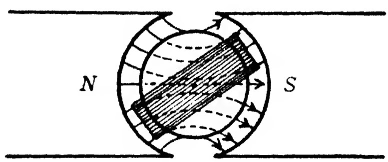


FIG. 154.—Radial magnetic field due to presence of iron core.

$$\theta = \frac{IHAN}{10T_o}. \tag{160}$$

Since for angles of 5° or less the cosine does not differ materially from unity, this factor is negligible when dealing with deflections of but a few degrees, even though the field is not radial. When used in connection with bridge circuits where a null method of adjustment is employed, a galvanometer having plane pole pieces is entirely satisfactory.

If Eq. (160) is written in the form

$$\frac{I}{\theta} = \frac{10T_o}{HAN\cos\theta} \tag{161}$$

we have an expression giving what is commonly designated as the **current sensitivity** of a galvanometer. In the ratio I/θ (current sensitivity) the current I is commonly expressed in microamperes and θ in terms of the deflection on a scale located 1 m from the galvanometer mirror. It is accordingly the custom to designate the current sensitivity of a galvanometer as the current (in microamperes) which will cause a deflection of 1 mm on a scale 1 m distant. In the case of most galvanometers having a fixed scale, the scale distance is ½ m; the standard sensitivity deflection would therefore be ½ mm. For instance, a certain form of wall galvanometer widely used in electrical measurements, and known as type P, has a current sensitivity ranging from 0.0002 μa to 0.014 μa, depending upon the number of turns in the coil used. Other high-sensitivity galvanometers of the reflecting type are available which have a sensitivity as high as 0.00001 μa, or 10^{-12} amp.

It was formerly the custom to rate a galvanometer in terms of **megohm sensitivity.** By this is meant the resistance in megohms that must be placed in series with the galvanometer winding in order that the standard deflection shall obtain when a potential difference of 1 volt is applied to the terminals. If the galvanometer resistance is neglected, the megohm sensitivity is numerically equal to the reciprocal of the current sensitivity. As implied above, this method of expressing sensitivity is rapidly going out of use; but models are still to be found bearing such a designation.

Still another method of rating galvanometer sensitivity involves the voltage which must be applied to the terminals, in series with critical-damping resistance (discussed below), in order to produce the standard deflection. This is known as the **voltage sensitivity,** and is expressed in terms of microvolts (μv). Reflecting galvanometers are made having voltage sensitivities ranging from 0.5 μv to 0.05 μv.

In making electrical measurements one may encounter circumstances which make it desirable to have an instrument whose deflection will indicate **quantity** rather than current magnitudes. By proper design, a galvanometer may be given such characteristics that it will perform this function. If, for example, one wishes to measure the quantity of electricity discharged from a condenser, it is desirable that the period of the galvanometer be made great as compared with the time required for the discharge to take place.[1] (By period is meant the time required for the swinging coil to describe a complete excursion.)

The period can be made relatively great (several seconds) by using a coil whose moment of inertia is large and by reducing both the mechanical

[1] The reasons for this requirement are discussed in any standard work on electrical measurements.

and electrical damping to a minimum. Such a galvanometer becomes, in effect, a coulombmeter, and is referred to as a **ballistic galvanometer.** The term **ballistic sensitivity** is used to indicate the quantity of electricity that must be discharged through such a galvanometer in order to bring about the standard deflection. Ballistic galvanometers are available having sensitivites ranging from 0.0003 μc to 0.01 μc.

In using a galvanometer to measure a quantity of electricity the first swing of the coil system is noted. The quantity causing the deflection is proportional to the resulting deflection, in conformity with the relation

$$Q = k\theta, \tag{162}$$

where θ is the deflection expressed in radians, and k a constant depending for its value upon the current sensitivity and the damping.

In addition to the sensitivity and the period of a galvanometer a third working constant should be noted. Reference is made to what is known as the **external critical-damping resistance.** By this is meant the value of the external resistance in the galvanometer circuit that will produce critical damping. And by critical damping is meant a condition such that the coil does not "overshoot" its quiescent reading when a current pulse is passed through the winding. When critically damped, the coil does not oscillate after having undergone a deflection. By means to be discussed later, it is possible to make a galvanometer "dead beat," as we say, even when on open circuit. For many uses to which a galvanometer is put, much time may be wasted in waiting for the coil to come to rest unless a condition of critical damping obtains. The value of the external resistance that produces critical damping ranges from a few ohms up to 100,000 ohms, depending upon the design of the instrument. In selecting a galvanometer for a particular use it therefore becomes necessary to consider all three characteristics, viz., **sensitivity, period, and external critical-damping resistance.**

In Fig. 155 may be seen illustrations of two widely used types of D'Arsonval galvanometers. The instrument shown in Fig. 156 is a portable D'Arsonval unit. In the portable model the suspension is short, and the coil carries a light pointer rather than a mirror, thus obviating the use of a telescope and scale. Such instruments have a current sensitivity of the order of 1 μa on their own scale; they are used chiefly as null indicators in bridge or potentiometer circuits.

The reader who desires to pursue the subject of the theory of the D'Arsonval type of instrument further will find the papers of Dr. Frank Wenner of the Bureau of Standards and of Dr. P. E. Klopsteg of Northwestern University of special interest. A booklet entitled "Notes on Moving-coil Galvanometers," published by the Leeds and Northrup

Company, also contains a good resume of galvanometer theory, together with a bibliography on the moving-coil galvanometer.

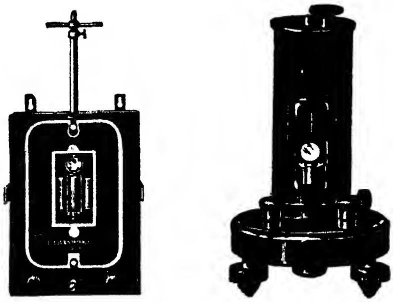


FIG. 155.—Modern laboratory types of D'Arsonval galvanometers. Wall type shown at the left; high-sensitivity form at the right. (*Leeds & Northrup Co.*)

FIG. 156.—Portable D'Arsonval galvanometer. (*Leeds & Northrup Co.*)

120. Vibration Galvanometer. In dealing with alternating currents it becomes necessary to measure capacitance, inductance, and phase angle (Sec. 159). Such determinations are usually made by null methods. A special form of detector for use in such measurements is known as the vibration galvanometer. This instrument consists of a solid coil of wire suspended between two pole pieces as shown in the illustration appearing as Fig. 157. The mass of the coil, together with the length and tension of the suspension, is so chosen as to give the moving system a natural mechanical period approximately equal to the electrical period of the current it is desired to detect. Provision is made whereby the mechanical period of the coil may be varied slightly. If and when an alternating current, say of 60 cycles, flows through the coil system it will, due to electromechanical resonance, vibrate through a small angle. If, then, a narrow band of light is caused to fall on the mirror attached to the coil the vibration of the coil will give rise to a widening of the reflected band as it appears on a suitable scale. Alternating currents whose frequencies differ materially from the natural frequency of the coil system will produce little if any effect. This is clearly shown by the graph appearing in Fig. 158. Galvanometers of this type are available having current

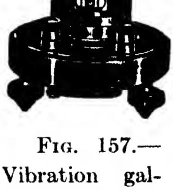


FIG. 157.—Vibration galvanometer. (*Leeds & Northrup Co.*)

sensitivities varying from 0.025 μa to 5 μa, when operating on 60-cycle current, and voltage sensitivities of 2 μv to 500 μv.

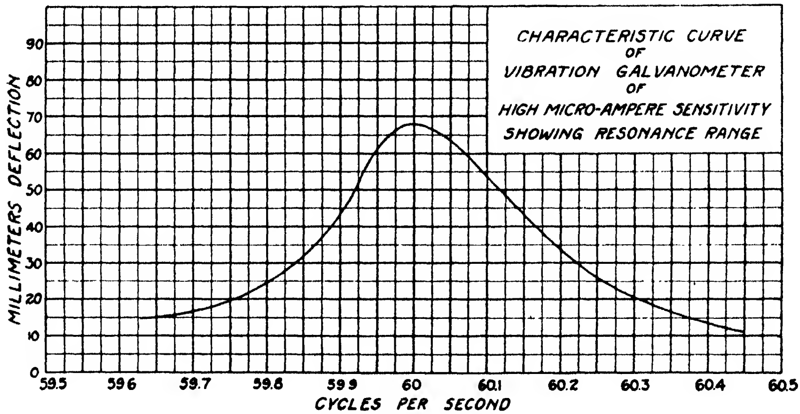


FIG. 158.—Relation between frequency and deflection in the case of a vibration type of galvanometer. (*Leeds & Northrup Co.*)

121. String Galvanometer. Another type of current-indicating instrument, having restricted but important fields of application, is known as the **Einthoven galvanometer.** This instrument is constructed

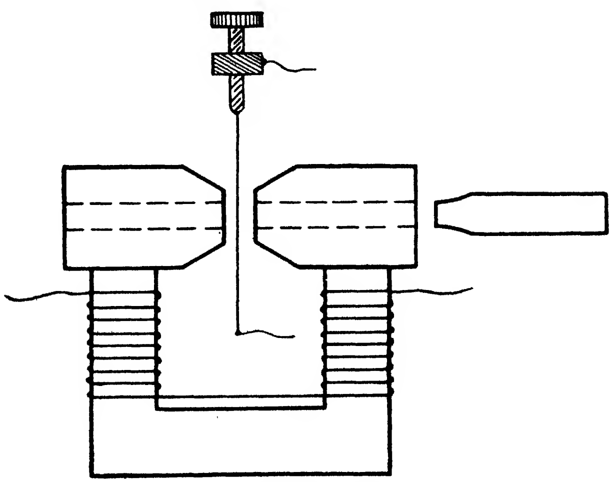


FIG. 159.—Einthoven galvanometer.

as diagrammatically sketched in Fig. 159. Between the poles of an electromagnet is stretched a fine fiber, usually made of quartz, which is rendered conducting by means of an extremely thin coating of silver, gold, or platinum. The current to be detected or measured is caused

to pass through the fiber element. The magnetic field due to this current reacts with the intense field due to the pole pieces of the electromagnet with the result that the fiber is slightly deflected at its middle point in a direction at right angles to the field and to the direction of the current. The movement of the fiber is observed through a hole in the pole pieces by means of a microscope equipped with a micrometer eyepiece. The sensitivity of the Einthoven type of galvanometer is of the same order as that of the D'Arsonval unit. In some applications of the Einthoven instrument the image of the fiber is projected, by means of a suitable optical system, on to a screen or photographic film. When observed

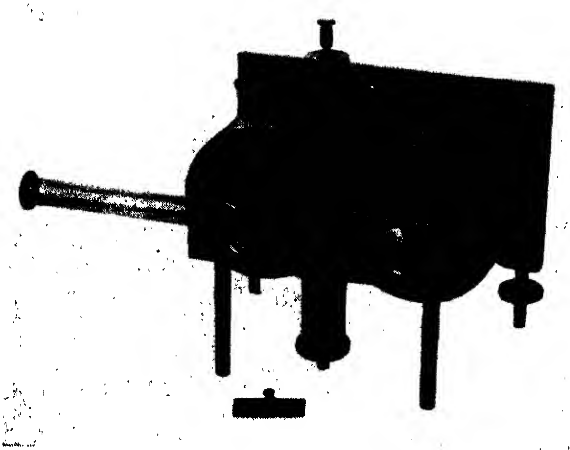


FIG. 160.—Laboratory form of Einthoven galvanometer.

by the latter means, this type of galvanometer may be used in the study of variable and alternating currents of extremely small magnitude. From the nature of its construction, the string galvanometer is quick acting and is also dead beat—two very important features in connection with certain applications. This type of instrument has been extensively used in several special fields of study, particularly in connection with medical investigations involving the use of the electrocardiograph, and also as a recording device in research in the fields of radioactivity and cosmic rays. Figure 160 shows a recent model of this type of galvanometer.

122. Ammeters and Voltmeters. The D'Arsonval type of indicating instrument may be made even more portable than the model shown in Fig. 156. This is accomplished by mounting the coil on pivots moving in jewel bearings. The current is led into and out of the winding through helical springs which also serve as control elements. A pointer attached to the coil system is free to move over a circular scale several

inches in length. The mechanical relation of these components may be seen in Fig. 161. The resistance of the coil winding is of the order of 5 ohms, and full-scale deflection is produced by a current of, say, 50 ma. In fact, such an instrument is actually a milliammeter. Such a unit can

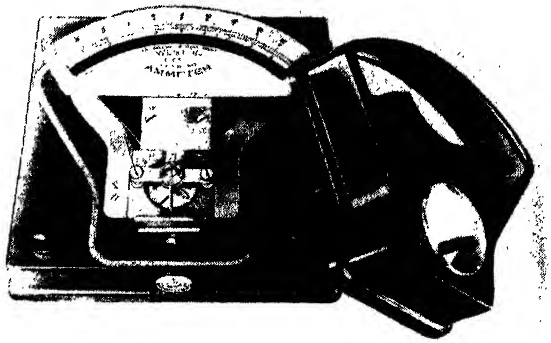


FIG. 161.—Interior view of D'Arsonval type of portable indicating meter. (*Weston Instrument Co.*)

be converted into an ammeter by the use of suitable shunts, as outlined in Sec. 71. The resistance of such shunts is so chosen that the currents causing full-scale deflection will be some convenient multiple of one another, as given by Eq. (100). Thus a single milliammeter may, with suitable shunts, function as several ammeters. In the case of low-range

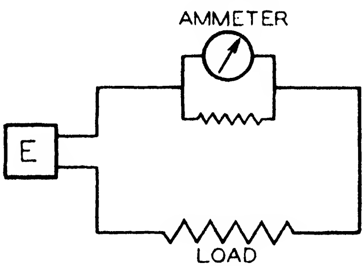


FIG. 162.—Showing manner of connecting a shunt to an ammeter.

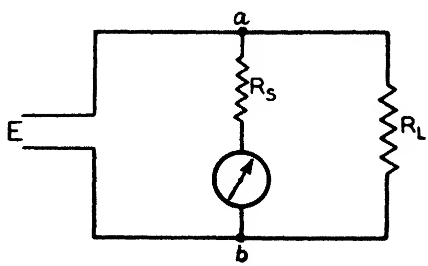


FIG. 163.—Method of connecting a voltmeter into a circuit. Note the series resistor.

ammeters, say 0 to 10 amp, the shunts are commonly housed within the body of the instrument. For the higher ranges, outside shunts are provided. An ammeter is connected in series with the source of potential difference and the load resistance, as indicated in Fig. 162.

A **voltmeter** is used to determine the potential difference **between two points in a circuit.** Can a milliammeter be made to function as a poten-

tial measuring device? The answer is in the affirmative; and this end is attained by inserting a high resistance in series with the instrument winding. This resistance should have a magnitude such that the current through the meter will not exceed the value necessary to produce full-scale deflection. For instance, if 50 ma will give a full-scale reading, and the internal resistance of the meter is 5 ohms, what external series resistor would be required if the meter was to be connected across a 120-volt circuit as sketched in Fig. 163? By Ohm's law we have

$$R = \frac{120}{0.05} = 2{,}400 \text{ ohms}$$

as the total resistance required between a and b in order that the current through the meter shall not exceed 50 ma. Since the internal resistance of the instrument is 5 ohms, the value of R_s in this case should be 2,395 ohms. If the potential difference between a and b exceeds 120 volts, the series resistance would have to be correspondingly greater. Under the circumstances just indicated the deflection of the instrument would be proportional to the potential difference applied to its terminals (including the series resistor), and may therefore be calibrated to read directly in volts. If the voltmeter is of the multiple-range type it will contain a series resistor for each range, the several resistors being connected in turn into the circuit by means of a plug or a multiple-point switch. Since the resistors carry only a few milliamperes, the wire can be small in diameter, and is commonly wound on a thin strip of insulating material located within the body of the instrument. It is to be noted that, in the particular case cited as an example, if the milliammeter had been connected directly across the 120-volt line the current would have been 24 amp, with the result that the wire constituting the winding would have melted.

123. Law of the Magnetic Circuit. We shall next consider the behavior of magnetic materials when placed in an electromagnetic field. In Sec. 30 the relations between magnetic field intensity H, flux density B, and permeability μ were discussed. In Sec. 117 of this chapter expressions relating current and field strength were developed. Both of these sections should now be reviewed.

In considering electromagnetism it should be observed that we are dealing with what may be called a magnetic circuit corresponding roughly to the electric circuit. In the magnetic case the lines of force are analogous to the paths of the electrons in the case of the current. There is, however, one marked difference between an electric and a magnetic circuit. An electric circuit may be "opened," so that the electrons will not flow, but a **magnetic circuit is always closed.** Lines of force always form

closed paths. For instance, in the case illustrated in Fig. 164, the magnetic circuit consists of the horseshoe magnet NMS, the two air gaps G and G', and the iron bar D. If the bar D was not there the lines of force would pass through the air between the N and S poles. There is a law of the magnetic circuit which corresponds to Ohm's law in an electrical circuit. Such a law has been known in a general way for some 200 years—possibly from the time of Euler, in 1761—but Rowland, in 1873, was the first to give definite form to this relation. In 1883 Bosanquet introduced the term **magnetomotive force** (mmf), corresponding to emf in the electric circuit. (The letter $\mathfrak{F}$ is sometimes used to designate magnetomotive force.) The law usually takes the form,

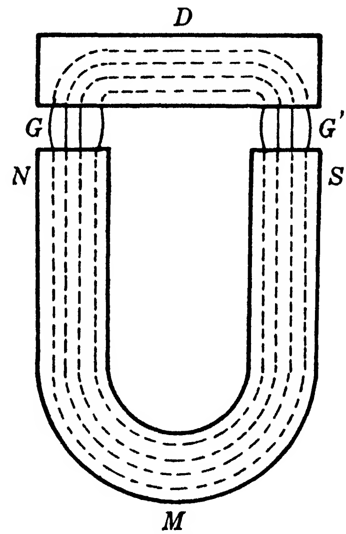


FIG. 164.—A magnetic circuit.

$$\text{Magnetic flux} = \frac{\text{magnetomotive force}}{\text{reluctance}}.$$

Expressing this in symbols we have

$$\Phi = \frac{\text{mmf}}{\mathfrak{R}} = \frac{\mathfrak{F}}{\mathfrak{R}}. \qquad (163)$$

The resemblance of this relation to Ohm's law is obvious. It will be recalled that the work done by a battery, when unit quantity of electricity is driven completely around the circuit, is a measure of the emf developed by the battery. Likewise the mmf established by the current flowing through a winding is numerically equal to the work done in transporting a fictitious unit pole once around the magnetic circuit. In general terms, the above statement might be written

$$\text{Work} = \text{mmf} = \oint H\,dl. \qquad (164)$$

In order to derive a working expression for mmf, let us assume that we have a toroidal winding as shown in Fig. 150, the length of the magnetic path being l. It has been shown [Sec. 117, Eq. (149)] that the field intensity in such a case is given by the relation

$$H = 4\pi nI,$$

where n is the number of turns per unit length and I the current in abamperes. Multiplying the above equation by l we have

$$Hl = 4\pi nIl.$$

But Hl would be the work done in transporting unit test pole around the magnetic circuit. Therefore, by definition,

$$\text{mmf} = 4\pi n Il.$$

For convenience in evaluating, we may let $N = nl$, where N is the **total** number of turns in the winding. Then

$$\text{mmf} = 4\pi NI = \mathfrak{F}. \tag{165}$$

The term NI is frequently spoken of as "ampere turns." It is evident that a given value of mmf may be produced by making the number of turns in the magnetizing coil relatively large and the current small, or by utilizing a comparatively large current and a few turns in the winding. In any given case, circumstances will determine the relative values of N and I. The unit of mmf is the **gilbert.** When one erg of work is done in moving a unit magnetic pole once around a magnetic circuit, the mmf is said to be one gilbert. If I is to be expressed in amperes, the above relation becomes

$$\text{mmf} = \frac{4\pi NI}{10} = 1.256NI. \tag{166}$$

Dealing next with the denominator of Eq. (163), it may be said that **reluctance** in the magnetic circuit corresponds, in a general way, to resistance in the electric circuit. In general terms it may be said that the term reluctance signifies the resistance which any substance offers to the setting up of magnetic flux. The magnitude of the reluctance is a function of the length of the magnetic path, the cross-sectional area, and the permeability. These factors are related thus,

$$\text{Reluctance} = \frac{l}{\mu a} = \mathcal{R}. \tag{167}$$

If a magnetic circuit is made up of several component parts, as shown in Fig. 164, the total reluctance of the circuit would be given by the relation

$$\mathcal{R} = \frac{l_1}{\mu_1 a_1} + \frac{l_2}{\mu_2 a_2} + \frac{l_3}{\mu_3 a_3} + \frac{l_4}{\mu_4 a_4},$$

where l_1 equals the length of the magnet NMS, l_2 equals the length of air gap G, l_3 equals the length of bar D, and l_4 equals the length of air gap G'. The several μ and a represent the permeability and cross-sectional area of the corresponding parts of the path.

From Eq. (38) it is seen that $\mu = B/H$; and we know that the ratio B/H is not constant. Hence it follows that **reluctance is not a constant.** In this important particular reluctance differs from resistance.

The unit of reluctance has no particular name. From the relation indicated in Eq. (163), it may be said that reluctance can be expressed in **gilberts per maxwell.**

If we now combine Eqs. (165) and (167), we have as the relation for any magnetic circuit

$$\Phi = \frac{4\pi NI}{l/\mu a}. \tag{168}$$

If the current is expressed in amperes the above equation becomes

$$\Phi = \frac{4\pi NI}{10(l/\mu a)} = \frac{1.256NI}{l/\mu a} \quad \text{maxwells.} \tag{169}$$

In the event that the reluctance of all parts of the magnetic circuit is not the same, our expression becomes

$$\Phi = \frac{1.256NI}{l_1/\mu_1 a_1 + l_2/\mu_2 a_2 + \cdots}. \tag{170}$$

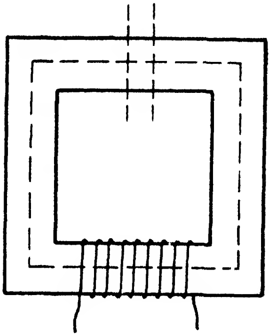


FIG. 165.—Effect of an air gap on the reluctance of a magnetic circuit.

The utility of the above expression may be illustrated by the examination of a specific case. Suppose we have a winding arranged to develop magnetic flux in an iron core, as shown in Fig. 165. Let us assume the mean length of the magnetic circuit, as shown by the dotted line, to be 60 cm. Let the cross section of the core be 5 × 5 cm. (*a*) What ampere turns will be required to develop a flux density of 15,000 gausses in the iron if the permeability at the specified flux density is 1,000; and (*b*) what ampere turns will be needed if an air gap 1 cm in length is cut from the core?

Combining Eqs. (41) and (169) we have, for (*a*)

$$NI = \frac{lB}{1.256\mu} = \frac{60 \times 15{,}000}{1.256 \times 1{,}000} = 717 \text{ ampere turns.}$$

For (*b*) we have

$$NI = \frac{B}{1.256}\left(\frac{l_1}{\mu_1} + \frac{l_2}{\mu_2}\right) = \frac{15{,}000}{1.256}\left(\frac{59}{1{,}000} + 1\right) = 12{,}570 \text{ ampere turns.}$$

If, for instance, the available current in the above two cases was 2 amp, the total number of actual turns of wire in the winding would be 358.5 and 6,285, respectively. It is thus evident that even a short air gap in a magnetic circuit greatly increases its reluctance. In designing electrical

machinery it therefore becomes important, in most cases, to reduce the air gap spaces to a minimum. In a few instances arrangements are made for adjusting the reluctance of a circuit by means of an adjustable air gap. In solving the second part of the above problem, we assumed that the cross section of the flux in the air gap is equal to the cross section of the iron core. Actually there is some spreading out of the flux ("fringing"). In practice the gap area is approximated by the use of empirical rules.

124. Magnetization Curves. In applying the relations embodied in Eqs. (165) to (170), inclusive, it becomes important to know in advance the magnetic properties of the ferromagnetic materials which are to be utilized. In Sec. 30 a preliminary survey was made of the relation that exists between the magnetic field strength H and the induction B that is established in a ferromagnetic material by the application of such a field. In practice magnetic material is commonly taken through a cyclic process of magnetization. It is necessary, therefore, to know how a given specimen of material will perform, magnetically, under those circumstances.

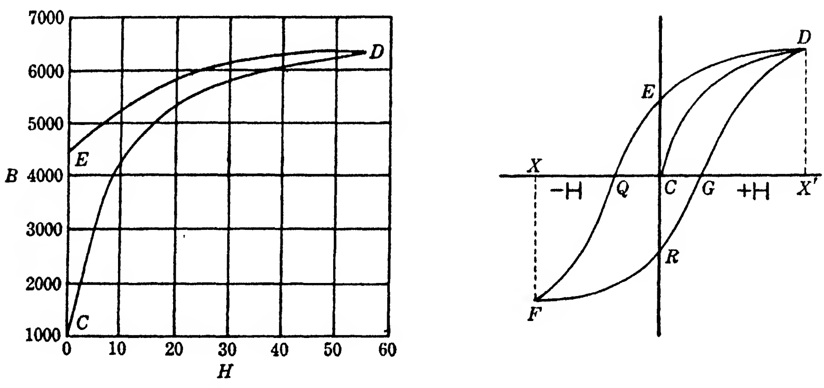


FIG. 166.—Magnetization curves and the hysteresis loop.

If a specimen of magnetic substance, such as iron, entirely free from magnetism, is subjected to a gradually increasing magnetizing field the flux in the material will increase quite rapidly at first and then more slowly; at length reaching a point beyond which an increase in the magnetizing field will not cause any increase in the flux. This is shown by the curve CD in Fig. 166a, the data for which were taken from an actual test. At something over 6,000 gausses this particular type of iron became saturated. If, after reaching the point D in the curve, the magnetizing current is gradually reduced to zero, the curve DE will result. It is evident that, though the magnetizing field has become zero,

a flux density of about 1,500 lines still exists in the iron. This flux is referred to as the residual induction or **remanence,** commonly designated by the symbol B_r.

Referring now to Fig. 166*b*, if the magnetizing current is reversed and gradually increased in value, the flux will continue to diminish, becoming zero when the magnetizing field has attained a reverse value represented by CQ. The magnetizing field that must be applied in the opposite direction in order to demagnetize a material after the initial magnetizing field has been reduced to zero is known as the **coercive force,** commonly designated by H_c.

If the magnetizing field be increased beyond Q, the flux will build up in the reverse direction as shown by QF on the curve, reaching saturation in the region of F. Reducing the magnetizing current from X to zero causes the flux to decrease in value, leaving a residual flux represented by CR. As a final step in the cycle of magnetization, the field may be increased from zero, in the original direction, with the result that the flux will vary as shown by that part of the curve marked RGD. When D has been reached the magnetic cycle has been completed.

An examination of the curve just traced discloses the fact the magnetic flux lags the magnetizing field throughout the entire cycle. This phenomenon is known as **hysteresis.** If our specimen of magnetic material were to be taken through subsequent magnetic cycles the flux would repeat the changes indicated by the loop $DQFGD$. The figure just described is known as the **hysteresis loop.** Owing to hysteresis the branch CD will not be described except when the specimen is entirely without magnetism at the beginning of the magnetic cycle. The difference between the magnitudes of the flux density due to the presence of the magnetic material and the flux density of the magnetizing field is referred to as the **intrinsic induction,** and is given numerically by the relation

$$B_i = B - H. \tag{171}$$

Figure 167 is reproduced from a photographically traced hysteresis curve for a sample of annealed iron wire.

How shall we account for the peculiar phenomenon just outlined? It is now assumed that the fundamental magnetic entity is the spinning electron. There is evidence that appears to indicate that when magnetization is modified, the direction or "sense" of the axial spin of certain of the electrons in the atom is changed, thus altering the over-all magnetic moment of the atom. The student who wishes to pursue this subject further should read a paper by R. M. Bozorth which appeared in *The Bell Telephone System Technical Journal,* January, 1940.

If and when a ferromagnetic material is subjected to a cyclic magnetizing field, at least a part of the energy involved in changing the direction of the electronic rotation manifests itself as heat. This results in the wasteful dissipation of energy. In other words, the material becomes heated and we have what is known as **hysteresis loss.** The nature of the magnetic material used, as well as the frequency of the cyclic changes and the maximum value of the induction, govern the magnitude of the heating effect, and hence the loss due to hysteresis. The magnitude of the loss resulting from hysteresis is proportional to the area of the hysteresis loop. If the scale of plotting is known, one may compute the hysteresis loss in watts by determining the area inclosed by the curve.

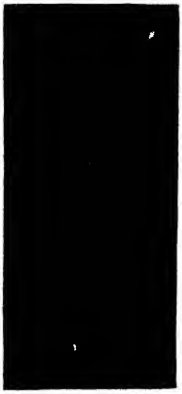


FIG. 167.—Experimentally recorded hysteresis loop for a sample of annealed iron wire.

Steinmetz developed an empirical formula for use in computing the loss due to hysteresis. It is

$$P = \eta V f B^{1.6}, \tag{172}$$

where P is the loss in watts, V the volume of the magnetic material in cubic centimeters, f the frequency in cycles per second, B the maximum flux density, and η is a constant the value of which depends upon the material being used for a core. The value of η varies from 0.0006, in the case of high-grade silicon steel, to 0.058 for hard tungsten steel. Equation (172) will not apply in those cases where the magnetizing field is not increased **continuously** from zero to its maximum value. (Why?)

It is interesting to note that, in general, the effect of impurities is to decrease the permeability and to increase the hysteresis loss; although there are certain exceptions to this statement. These exceptions are of considerable commercial importance. For example, the presence of a small percentage of silicon or aluminum increases the permeability and decreases the hysteresis loss. As a result of this, silicon steel (1 to 4 per cent silicon) is extensively used as the core material in the design of power transformers and dynamos. By subjecting pure iron to a special heat treatment *in vacuo*, Yensen has produced an iron having a permeability higher than that shown by the best iron previously obtainable and a hysteresis loss of only a third that of standard transformer steel. A magnetic alloy known as **perminvar** (iron, nickel, and cobalt) is said to be almost entirely free from hysteresis when worked at low flux density. This material also shows only slight coercive force.

Still another magnetic alloy, known as **permendur,** is composed of 50 per cent iron and 50 per cent cobalt. Its saturation value is extremely

high, being of the order of 25,000 gausses at 1,250 oersteds. This material is particularly useful in the design of d-c electromagnets.

In the field of communication engineering it is particularly important to have available a magnetic material that shows high permeability and low hysteresis loss in weak fields. A magnetic alloy known as **permalloy** possesses these desirable properties. One form of this material consists of approximately 80 per cent nickel and 20 per cent iron and is said to have a permeability 200 times as great as that of iron. Figure 168 shows the *B-H* curves for a specimen of this alloy as compared with a sample of high-grade silicon steel.

In the design of transformers that must function at frequencies of the order of 100,000 cycles/sec., it becomes necessary to have available, if

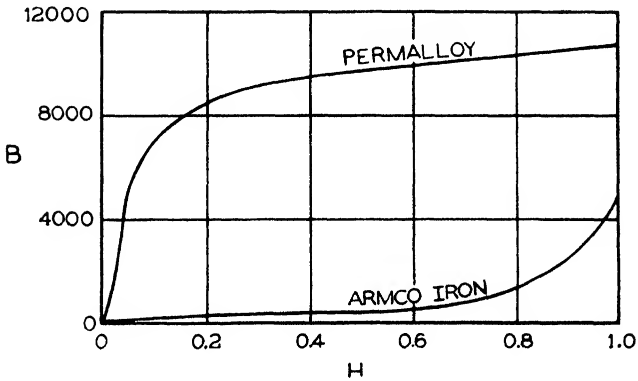


FIG. 168.— Magnetization curves of iron and permalloy.

possible, a core material whose permeability is approximately constant and which shows a negligible hysteresis loss.

It has been found that either iron or permalloy dust, mixed with such a material as resin and pressed into usable shape, answers this purpose quite satisfactorily. Such core material is now extensively used in radio receiving sets.

If a material is to be used for permanent magnets, it is important that its coercive force and its remanence be high. K. Honda, a Japanese metallurgist, has produced a cobalt-steel alloy which contains, in addition to the steel, about 35 per cent cobalt, 8 per cent tungsten, and 3 per cent chromium. This special steel is very resistant to demagnetization, its coercive force being three or four times that of tungsten steel. The Honda alloy is utilized to some extent as the material for magnets that are to be used in devices requiring great permanency of the magnetic field.

Recently another important magnetic alloy has been developed by Nesbitt and Kelsall of the Bell Telephone Laboratories. Known as **vicalloy,** it is composed of cobalt, vanadium, and iron. This alloy is said to have a greater coercive force than any other commercial material. It can be drawn into fine wire and rolled into a thin tape. Because of its magnetic and mechanical properties it is being utilized as a medium for the magnetic recording of voice and music (Sec. 251).

125. Barkhausen Effect and Magnetostriction. In dealing with the magnetic behavior of materials, one encounters several singular phenomena. If an experimental assembly is set up, as shown in Fig. 169, definite clicks will be heard in the telephone headset as the current is slowly increased in the magnetizing coil about the rod or bar of magnetic material. This would appear to indicate that the flux in the iron changes **by small discrete increments** rather than in a continuous manner. The phenomenon is called the **Barkhausen effect,** after its discoverer.

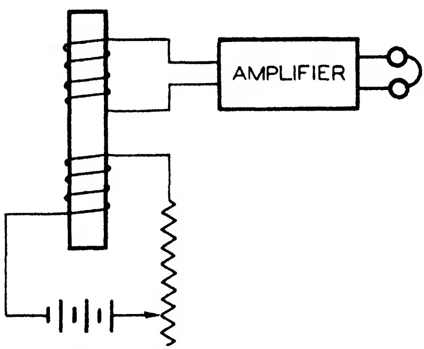


Fig. 169.—Plan for observing the Barkhausen effect.

A second magnetic phenomenon, discovered by Bidwell in 1886, is also of interest. Bidwell found that changes in the dimensions of bodies occur when they are magnetized. For instance, a wire or rod of magnetic material shows a measurable change in length when subjected to a magnetic field. If the applied field is of a variable nature, the magnetic material will expand and contract in response to the changing magnetizing force. The phenomenon of a change in one or more dimensions of a body when it undergoes magnetization is termed **magnetostriction.** Pure iron and steel exhibit only very feeble magnetostrictive effects. Pure nickel, however, shows marked magnetostrictive response. Certain alloys also exhibit this phenomenon. While the absolute change in dimensions due to magnetostriction is small, this effect is utilized in certain special electrical equipment for use in connection with the production of h-f alternating currents, and also in the production of supersonic vibrations.

The converse of the above-mentioned effect is also encountered. For instance, a nickel rod in a magnetic field shows an increase in magnetization when subjected to longitudinal compression and a decrease when tension obtains. A small strain may produce a marked effect in the magnetization. This reverse phenomenon is called the **Villari effect.**

It seems probable that whatever the final explanation of magnetism may turn out to be it will have to account for both the Barkhausen effect and magnetostriction. It is possible that an intensive study of these, and related phenomena, might yield valuable information as to the ultimate nature of magnetism.

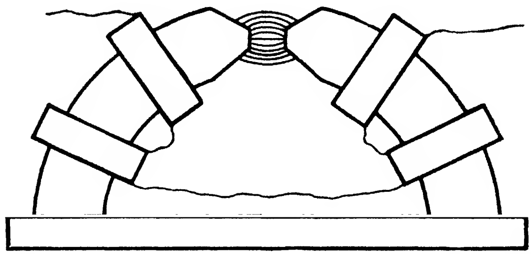


FIG. 170.—Arrangement for producing intense magnetic fields.

126. Electromagnets. Except in the case of certain electrical measuring instruments, electromagnets having soft-iron cores are utilized for the purpose of producing intense magnetic fields. Owing to the form of the magnetization curve it is not practicable to subject the iron used in transformers, dynamos, and special electromagnets to flux densities

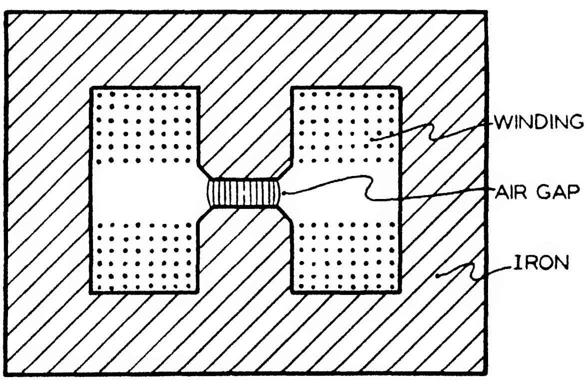


FIG. 171.—Yoke type of magnetic circuit.

greater than about 10,000 gausses. Hence, in order to secure the required number of lines, suitable cross-sectional areas must be provided, and the magnetic circuit kept as short as possible in order to hold the reluctance at a minimum value. Equation (170) is applicable in such cases. The magnetic circuits of dynamos and transformers will be considered in a later chapter.

Various special electromagnets are of great interest because of their use in connection with highly important research and development work. Very intense magnetic fields are often needed for such purposes. For instance, a concentrated, high-intensity field is essential in connection with the mass spectrograph (Sec. 207). Such electromagnets often take a form similar to that shown in Fig. 170. By shaping the pole tips in the form of a cone or a wedge highly concentrated fields may be produced.

If the flux is to be distributed over a relatively large pole-face area, as in the construction of a cyclotron or a betatron (Sec. 217), the double-yoke type of design is usually employed. The plan of such a magnetic circuit is diagrammatically indicated in Fig. 171. In such cases the pole faces may be several feet in diameter and the highest possible flux

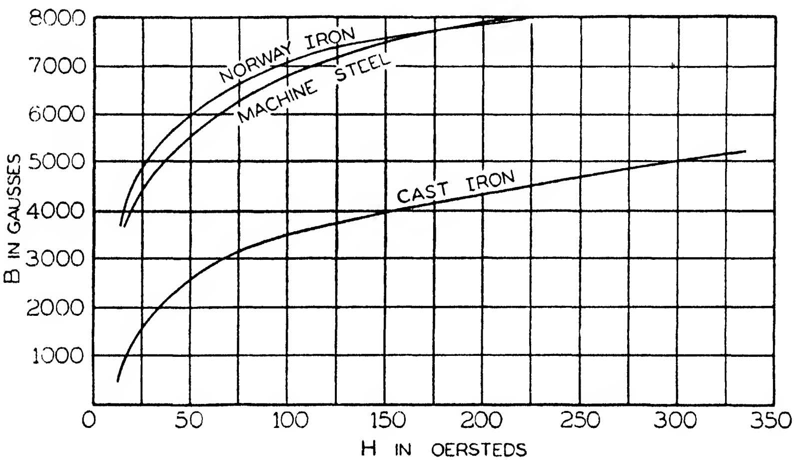


FIG. 172.—Magnetization curves.

density in the gap is needed. In order to develop the flux required in such cases, the iron involved may amount to two or three hundred tons and the field windings may consist of several thousand turns of heavy copper conductor, often in the form of tubing cooled by water or oil.

127. Magnetic Testing. From what has been said in the preceding section it is obviously of great importance, in connection with the design of all equipment that involves a magnetic circuit, to have available accurate data bearing on the magnetic properties of the material proposed for use. Hence the testing of magnetic materials forms an important part of electrical measurements. Magnetization and permeability curves corresponding to those shown in Figs. 172 and 173 are extensively used as aids in design calculations.

For example, if one is designing a transformer which is to be used in a telephone circuit, it is important to arrange conditions so that the iron will not be worked at or near the saturation point; otherwise speech distortion will result. In such a case as this the volume of iron and the number of turns in the windings would be so laid out that the core would be worked on the comparatively straight part of the *B-H* curve. In order to secure data from which to plot the *B-H* and μ-*H* curves, together with the complete hysteresis loop, various methods are followed. It is outside the scope of this volume to describe these methods in detail.

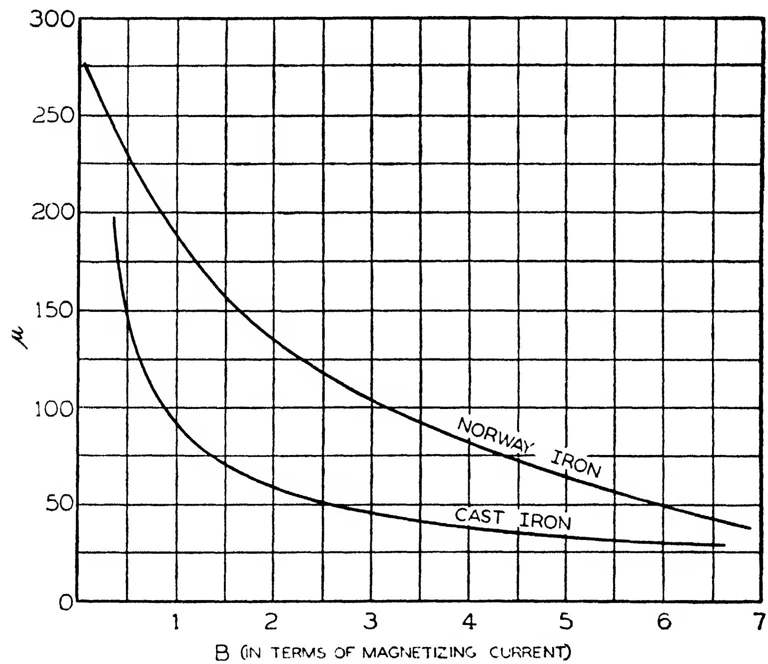


FIG. 173.—Permeability curves.

However, it may be said that the general process consists in first completely demagnetizing the sample of material and afterwards subjecting it to a magnetizing field the value of which is changed by definite steps. Means are provided whereby the resulting flux in the material can be directly or indirectly measured as the magnetizing field is changed. Such a measurement commonly involves a small "test coil" connected to some form of a ballistic galvanometer. This coil is placed about the sample under test. When the magnetic flux is changed in magnitude an emf will be developed in the winding constituting the test coil, and the resulting galvanometer deflection will be a function of the flux magnitude.

Portable direct-reading fluxmeters are now commercially available, such, for example, as the one illustrated in Fig. 174. For a complete treatment of the methods used in magnetic testing the student is referred to the two

FIG. 174.—Portable fluxmeter. (*Sensitive Research Instrument Co.*)

following publications of the U.S. Bureau of Standards: *Bulletin*, Vol. 6, p. 31, and "Magnetic Testing," *Circular* 17, or to any standard laboratory manual on electrical measurements.

PROBLEMS

1. What will be the magnetic field strength at a point 10 cm from a long wire carrying a current of 10 amp?

2. What will be the mechanical force experienced by a magnetic pole whose strength is 50 cgs units if placed at the point indicated in the above problem?

3. What will be the magnetic field intensity due to the current in a conductor that is in the form of a closely wound circular coil at a point on the axis 25 cm from the center of the coil. The coil has 20 turns, the mean radius of which is 10 cm. The current is 5 amp.

4. A pair of Helmholtz coils, having the specifications given in Prob. 3, and carrying a current of 5 amp, are supported 10 cm apart. What will be the field strength at a point on the axis of the coils midway between the windings?

5. A helical winding is 100 cm long and contains 500 turns. If a current of 5 amp flows in the winding, what will be the field strength near the center of the winding? Near the ends?

6. What will be the total flux if a bar of iron occupies the space within the winding of Prob. 5, assuming the permeability of the iron to be 500?

7. Find the ampere turns necessary to produce a flux density of 10,000 gausses in a closed ring or iron having the following specifications: Mean diameter of ring = 15 cm, diameter of iron core = 2.5 cm, and permeability = 500.

8. What ampere turns would be required to produce the same flux density if, in the iron of Prob. 7, the core had an air gap of 5 mm?

9. What will be the mechanical force experienced by a conductor at right angles to the direction of a uniform magnetic field whose intensity is 1,000 oersteds? The effective length of the conductor is 10 cm, and the current is 5 amp.

10. Show that the mechanical force exerted upon one another by two parallel conductors, carrying currents of I_1 and I_2, respectively, and located d centimeters apart, is given by the expression $2(I_1I_2/d)$ dynes per unit length.

11. A rectangular loop measures 16 × 8 cm, and is supported by, and is free to rotate about, an axis through its center and parallel to its greater dimension. If, in a field whose intensity is 1,000 oersteds, the plane of the coil makes an angle of 30° with the direction of the flux, what torque will be developed when a current of 5 amp flows in the conductor?

12. It is desired to use a certain milliammeter as a voltmeter. There are 100 scale divisions and each division represents 1 ma. If the internal resistance of the meter is 100 ohms, what must be the value of the resistance that must be inserted in series with the winding in order that full-scale deflection shall indicate 100 volts?

13. The iron core of a certain transformer has a volume of 2,050 cm³. The core material (silicon steel) has a coefficient of hystersis loss of 0.001. Compute the hysteresis loss in the core when the transformer is operated at a maximum flux density of 8,000 gausses and a frequency of 60 cycles/sec. What would be the core loss if the frequency were 1,000?

14. The hysteresis loop for a given sample of iron is plotted on centimeter cross section paper. On the Y-axis 1 cm represents 1,000 gausses, and on the X-axis 1 cm indicates 20 oersteds. A planimeter reading of the area enclosed by the complete loop gives 11.6 cm². From the foregoing data compute the energy loss in ergs per cubic centimeter per cycle.

CHAPTER XVI

ELECTROMAGNETIC INDUCTION

128. Induced Electromotive Force. On Nov. 24, 1831, 11 years after Oersted's basic discovery, Faraday read a paper before the Royal Society of London in which he made known his discovery that whenever the magnetic flux in the region of any conductor is caused to vary in intensity that conductor becomes the seat of a temporary emf. If the conductor in which the emf is developed forms a closed circuit, a current will result, and this current is referred to as an **induced current.** This discovery was of transcendent importance. It is doubtful whether any one event in all history has produced a greater effect on the material aspects of human society than has this observation by Faraday. Together with Oersted's findings, Faraday's disclosures form the basis of communication engineering and all electrical power generation and distribution.

Faraday, however, was not alone in the early work done in connection with what has come to be spoken of as electromagnetic induction. The American, Joseph Henry, also made valuable contributions to our knowledge of the fundamental facts in this field. Indeed, it is probable that the possibility of obtaining an electric current through the agency of a magnetic field was recognized independently by Henry at about the same time that Faraday was carrying on his induction experiments in England. The results of Henry's researches, however, were not made known until after Faraday had disclosed the results of his investigations. Hence the honor of the discovery quite properly goes to Faraday.

The facts discovered by Faraday and other early investigators, in connection with the production of an induced emf and the resulting currents, led to the formulation of laws that serve to relate the various observed phenomena. Emil Lenz presented a paper before the Academy of Sciences at St. Petersburg in 1833 in which he gave a simple law by means of which the direction of an induced emf may be predicted. This rule, known as **Lenz's law,** is to the effect that **the direction of an induced current is always such that by its electromagnetic action it tends to oppose the motion which gives rise to it.** The law is of great importance in many of the applications of electromagnetic induction, as we shall see in later sections.

In 1845, F. E. Neumann developed a quantitative expression embodying Faraday's experimental findings. Expressed in terms of the calculus,

Neumann's description of the relation existing between the cause and effect in electromagnetic induction takes the form

$$e = -\frac{d\Phi}{dt}, \tag{173}$$

where e represents the induced emf and $d\Phi/dt$ the time rate of change of magnetic flux. In other words, the magnitude of the induced emf depends upon the rate at which the magnetic flux changes in value in the region of the conductor in which the induction is taking place. The negative sign indicates that the induction takes place in conformity with Lenz's law. **The magnitude of the induced emf is independent of the particular method by which the magnetic field is caused to vary in intensity.** In dealing with electromagnetic induction it should also be emphasized that **the emf is what is induced.** If the circuit in which the induction takes place is not closed, no current will flow; but an induced emf may, however, exist.

Equation (173) gives the **instantaneous** value of the induced emf. The **average** value is given by the ratio of the total flux change to the time involved in that change, or

$$E = \frac{\Phi}{t} \qquad \text{emu (abvolts).} \tag{174}$$

Thus if the average flux change is at the rate of one line of force per second, the average induced emf will be 1 emu. Obviously this is an extremely small emf. The International Electrical Congress, which met in Paris in 1881, decided to call the practical unit of potential difference and emf the **volt** and to fix its value at 10^8 emu (abvolts). This is a value of emf approximately equal to that developed by the Daniell cell, which, at that time, was considered to be a dependable standard of emf. On the basis of the above definition the emf in volts would be given by

$$E = \frac{\Phi}{10^8 t}. \tag{175}$$

129. Emf Developed in a Linear Conductor Due to Its Movement in a Magnetic Field. Not only is it possible to develop an emf in a conductor by causing a magnetic field to vary in intensity in the neighborhood of the conductor, but an emf may also be established in a conductor by bringing about relative motion between the conductor and the field. The fundamental relation embodied in Eq. (174) is applicable to this case also. We shall find it useful to have available an expression giving the value of the emf induced in a linear conductor when being moved at constant linear velocity through a magnetic field.

In Fig. 175, suppose that a straight conductor is moved perpendicularly to the direction of the flux in the time t, through a uniform field from the position ab to the position $a'b'$. Let the length of the conductor be designated by l, and the linear displacement by s. From Eq. (41) we have

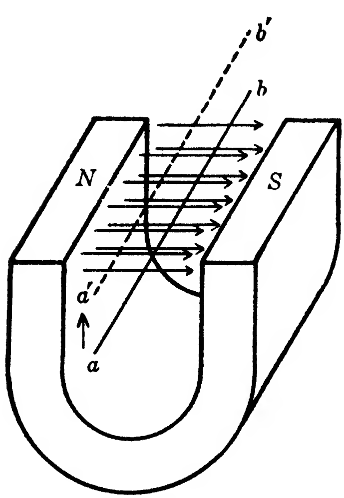


Fig. 175.—Emf developed in a conductor due to its movement through a magnetic field.

$$\Phi = AB,$$

where Φ is the total flux, A the area, and B the flux density. The area swept out by the movement of the conductor will be ls. Hence

$$\Phi = lsB.$$

Dividing by t we get

$$\frac{\Phi}{t} = lB\frac{s}{t}.$$

But s/t = velocity = v. Hence we may write

$$E = lBv \qquad \text{abvolts}, \tag{176}$$

where l is the active length of the conductor in centimeters, B the flux density of the inducing field, and v the velocity of displacement in centimeters per second. If E is to be expressed in volts the above relation becomes

$$E = \frac{Bvl}{10^8}. \tag{177}$$

It should be noted that the same result would have been obtained had the conductor remained stationary and the magnet been moved.

130. Emf Induced in a Rotating Coil. The results of the last section may be extended to cover the case of a conductor having the form of a rectangular loop, as shown in Fig. 176*a*. Assume that the loop *abcd* is rotated in a clockwise direction about the axis XX'. It is evident that this case differs from the previous situation in two respects. First, we have **two** active conductors *ab* and *cd* connected in series; and secondly, both of these conductors, because of their angular displacement about the axis XX', will "cut" flux at a changing rate.

If we consider the side *ab* as it moves through some given position, such as that shown in Fig. 176*a*, the emf developed in this part of the conductor will be in a direction away from the reader. At the same time the emf produced in the side *dc* will be directed from *c* toward *d*. The

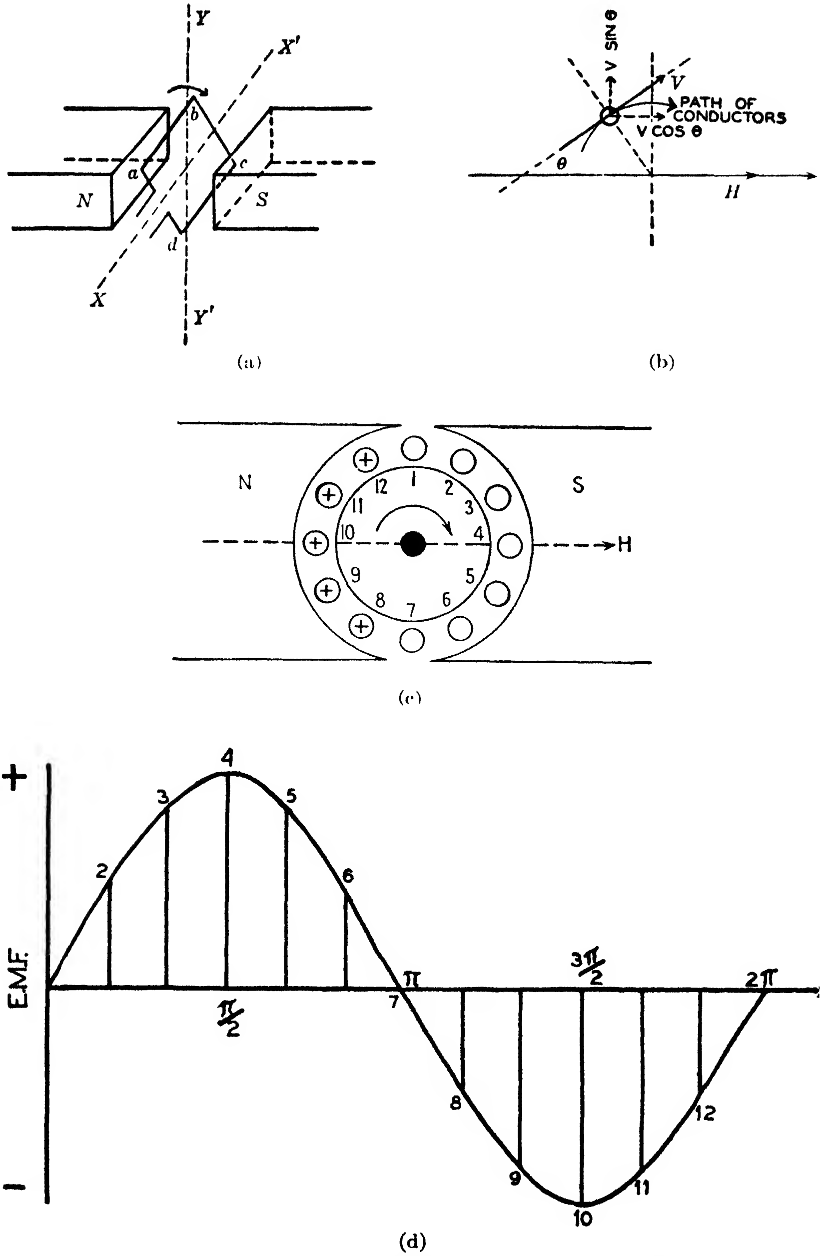


FIG. 176.—Emf developed in a rectangular coil rotating in a magnetic field.

two emf are, therefore, additive. Hence, by applying Eq. (177), we would have as the average emf

$$E = \frac{2Bvl}{10^8} \qquad \text{volts.} \tag{i}$$

Since the end sections *bc* and *ad* of the loop move in a direction parallel to the flux, no emf will be produced in these parts; and hence they will contribute nothing to the emf generated by the loop as a whole.

While the linear velocity of the loop may be constant, its **effective** velocity is variable. Referring to Fig. 176*b*, it will be seen that at any instant the actual velocity vector makes an angle θ with the direction of the flux, and that the **component** of the velocity **perpendicular to the direction of the flux** will be $v \sin \theta$. This component, $v \sin \theta$, is the effective velocity of the conductors with respect to the flux. From the geometry of the case, θ is also the angle between a vertical plane through the axis of rotation and the instantaneous position of the plane of the coil; and it is this angle that is meant whenever the angular displacement of the coil is referred to. Replacing v in (i) by its effective component $v \sin \theta$, we have

$$e = \frac{2Blv \sin \theta}{10^8}, \tag{ii}$$

where e is the value of the emf **at any instant.**

It will be more convenient to have the velocity of the conductors expressed in angular measure. To do this we make use of the relation

$$v = \omega r,$$

where ω is the angular velocity and r the radius of the rotating coil. But

$$\omega r = 2\pi R r,$$

where R is the number of revolutions of the coil per second. Hence we may write

$$e = \frac{4\pi BRrl}{10^8} \sin \theta. \tag{iii}$$

It is evident that $2rl = A$, the area of the coil; and since $\Phi = BA$, (iii) becomes

$$e = \frac{2\pi R\Phi}{10^8} \sin \theta.$$

If, as is commonly the case in practice, the coil consists of several turns connected in series, each loop will contribute its own emf, and hence

the total emf generated in the winding would be

$$e = \frac{2\pi NR\Phi}{10^8} \sin \theta, \tag{178}$$

where N is the number of turns in the coil.

From the last equation it may be seen that the maximum value of the emf will obtain when $\sin \theta = 1$, which will be when θ is 90°, *i.e.*, when the plane of the coil is parallel to the direction of the flux. In that position a given conductor will be moving at right angles to the flux; therefore $d\Phi/dt$ will be a maximum. The minimum value of the emf will occur when $\theta = 0$, which occurs when the plane of the coil is normal to the flux. In this position the conductors will be moving parallel to the flux and hence $d\Phi/dt = 0$. At intermediate positions the emf will have values between zero and the maximum corresponding to the various values of $\sin \theta$. Since $\theta = \omega t$ it is seen that the emf developed in the loop is a periodic function of the time, *i.e.*, an alternating emf will be produced in the winding. In passing it is to be noted that when the emf is a maximum, the direction of the magnetic field, the direction of the emf, and the direction of the motion **are mutually perpendicular to one another.**

There is one other significant aspect of the case that should also be noted. Figure 176*c* is intended to represent a number of the positions that one of the conductors, say *ab* in Fig. 176*a*, will assume during one complete revolution of the coil. For convenience these positions are spaced 30° apart. When the conductor is in position 1, the emf will be zero, as noted above. As it moves through positions 2 and 3 to 4, the emf will increase in value and its direction will be toward the reader. As it passes on through 5 and 6 to 7, the emf will decrease, becoming zero again at 7. From position 7 to 10 the emf will again increase in value, but, since for this particular conductor the direction of the flux is in effect reversed, **the emf will be reversed also** and will accordingly be directed from *a* toward *b*. From 10 to 1 the emf will again decrease, becoming zero at 1. It will thus be seen that the emf is periodic, or alternating, in character. If these angular positions and the corresponding emf values are plotted, the result will be the graph shown in Fig. 176*d*. This curve will represent one complete **cycle** of operations. The time required for the completion of a cycle is known as the **period.** The **frequency** would be the number of cycles which occur per second.

131. Foucault or Eddy Currents. Variations in magnetic flux give rise to induced emf and resulting currents not only in linear conductors, such as wires, but also in conductors having the form of plates or sheets. Before Faraday's discovery of magnetic induction, Gambey, in 1824, noticed that the motion of a magnet suspended and set into oscillation

was rapidly damped when a plate of copper was held just beneath the magnet. At about the same time Arago observed that when a pivoted magnetic needle is supported directly over a rotating copper disk the magnet is caused to rotate with the disk. Faraday explained these phenomena on the basis of electromagnetic induction. In the latter case the relative motion of the disk in the field of the magnet gives rise to emf in the body of the metal disk, which, in turn, sets up currents that circulate in closed paths. The magnetic fields resulting from these induced currents are, by Lenz's law, of such direction as to oppose the motion which gives rise to them; hence the mechanical effect upon the suspended magnet. The magnitude of the emf induced in the body of the conducting material may be small, but, since the resistance of the conductor may be, and often is, extremely low, large current values may obtain. As a result, these so-called **eddy currents** give rise to strong magnetic fields. These magnetic fields, reacting with the field that gave rise to them, tend to produce marked mechanical effects. A number of important commercial applications of the "magnetic drag" effect have been made.

For instance, in one widely used type of automobile speedometer, an annular magnet is caused to rotate by means of a spiral driving shaft in juxtaposition to an aluminum disk, the motion of the latter being controlled by means of a spiral spring. As the magnet revolves, the eddy currents in the disk cause it to turn and thus indicate the speed. The calibration marks appear on the edge of the disk.

Another application of this principle is found in the control mechanism which forms a part of watt-hour meters (Sec. 176). Certain types of these meters consist of a set of fixed coils and a rotating coil, the latter acting, in effect, as the armature of a motor. To control the speed of the rotating element in such meters a metallic disk is fastened to the shaft of the rotor. One or more permanent magnets are fastened with poles close to the surface of the disk. As the armature rotates the currents induced in the disk tend to oppose the motion of the rotor, thus acting as a stabilizing mechanical load. The speed of the rotor can be adjusted by altering the position of the magnets. This control disk is usually visible from the outside of such meters, and may be seen rotating whenever current is passing through the meter.

Still another application of magnetic damping is to be found in the aluminum coil form that supports the windings of d-c ammeters and voltmeters. As the coil of the moving system rotates, in response to the electromechanical torque, eddy currents are set up in the metallic coil frame, and the resulting magnetic field tends to dampen the motion of the coil and thus make the instrument "dead beat."

Perhaps the most important commercial application of eddy currents is to be found in the induction motor. By means of a certain arrangement of polyphase a-c circuits, to be studied later, it is possible to produce a rotating magnetic field. If a rotor consisting of short-circuited low-resistance conductors is placed within the region of such a rotating field large eddy currents will be developed, with the result that the conductors forming the rotor will experience a torque tending to produce rotation, thus converting electrical energy into mechanical energy.

The foregoing are illustrations of some conditions under which eddy currents serve a **useful** purpose. There are, however, circumstances under which **such currents are very undesirable.** It appears to have been first observed by Foucault that these eddy currents result in heat, and consequently give rise to a loss of energy. When a metal disk is rotated rapidly between the poles of a strong electromagnet, its temperature is decidedly raised; the I^2R law operates as usual.

If the armature core of a dynamo was made of solid iron, as sketched in Fig. 177*a*, and was rotating, emf would be induced in the surface layers of the iron, and the resulting eddy currents would heat the core.

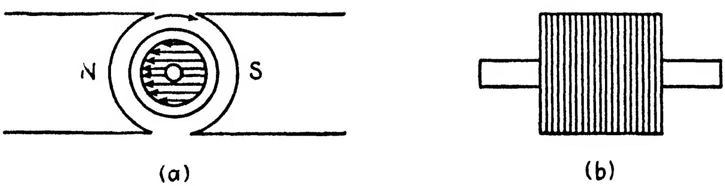


Fig. 177.—Reduction of eddy currents in the armature core of a dynamo by the use of thin laminations.

Since, in this case, these currents would serve no useful purpose, their existence means a wasteful dissipation of energy. If, however, the armature core were to be built up of thin insulated laminations, as shown in Fig. 177*b*, the resistance offered by the metal as a whole to these Foucault currents would be materially increased; indeed, the eddy currents may be almost completely suppressed by such a method of construction. The plane of the laminations must, of course, be at right angles to the direction of the Foucault currents. (Why?) The thinner the laminations, and the better they are insulated from one another, the more effective are they in preventing these wasteful currents.

In transformers (Sec. 146) the varying flux set up by the alternating current flowing in the windings tends to give rise to Foucault currents in the core, thus developing heat. By building up the core of thin insulated laminations placed as shown in the projection in Fig. 178, these

undesirable currents are largely prevented, and the so-called core losses[1] materially reduced. The core laminations are usually insulated from one another by being varnished or enameled.

It may be shown[2] that the energy dissipation due to eddy currents developed in thin sheets can be computed by the relation

$$P_e = \frac{\pi^2 t^2 f^2 B^2_{max}}{6\rho 10^{16}}, \tag{179}$$

where t is the thickness of the laminae in centimeters, f the frequency in cycles per second, B_{max} the maximum flux density in lines per square centimeter, and ρ the specific resistance of the material constituting the sheets. P_e will be the loss in watts per cubic centimeter.

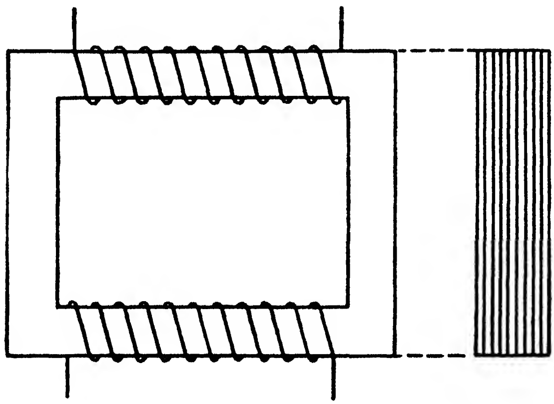


FIG. 178.—Use of laminations in the core of a transformer for the purpose of reducing eddy-current loss.

An examination of Eq. (179) discloses several important facts. In particular it is to be noted that the eddy-current loss varies **inversely as the specific resistance of the material involved.** In Sec. 124 it was pointed out that certain magnetic alloys have comparatively low hysteresis coefficients. It is a fortunate circumstance that such alloys show relatively high specific resistance; in some cases eight times that of ordinary iron. Thus both the hysteresis and the eddy-current losses are reduced by the use of such alloys.

Further, the equation tells us that the eddy loss varies **directly as the square of the frequency** of the magnetizing field. Hence the loss increases rapidly at the higher frequencies, especially those encountered in con-

[1] The core or "iron loss" in electrical equipment involving varying magnetic fields consists of two parts, viz., eddy-current losses and losses due to hysteresis.

[2] See "Alternating Current Phenomena," Chap. XLV, by Steinmetz, or "Alternating Currents," by A. E. Clayton, p. 127.

nection with the use of audio transformers. It is also to be observed that the loss varies **directly as the square of the maximum flux density** at which the core material is being worked.

While the heating effect due to eddy currents may result in a wasteful dissipation of energy, such thermal effects may, under certain circumstances, be made to serve a useful purpose.

In Sec. 85 reference was made to a so-called induction furnace. Furnaces of that type make use of the heating due to eddy currents. The fact that the thermal effect varies as the square of the frequency is utilized when it is desired to develop heat in this manner. The Northrup furnace is operated at a frequency of the order of 50 M cycles. Quite recently the heating of metals by eddy currents operating at frequencies in the region of 50 megacycles is coming into extensive industrial use. Such a method of developing thermal energy is referred to as **high-frequency heating**; when applied to certain metallurgical processes, h-f heating possesses a number of important advantages over some of the older procedures. High-frequency heating also finds use as a therapeutic agent. Thus we see that Foucault currents, while they may be undesirable under certain circumstances, under other conditions may serve highly important and useful purposes. The methods by which h-f alternating currents, of the order indicated above, may be generated will be discussed in Chap. XXIX.

132. Emf of Self-induction. It has already been shown (Sec. 128) that a changing magnetic flux value will give rise to an induced emf in a circuit located in that field. It is also true that if a conductor in the form of a coil is carrying a varying current, the resulting variations in the flux will induce an emf **in the circuit itself.** This holds even though the coil consists of only **one** turn. In conformity with Lenz's law, if the current in the coil be **increasing** in value the self-induced emf will **oppose** the applied emf and hence retard the rise of the current in the circuit. Similarly, if the original current in the coil be **decreasing,** the self-induced emf will be in the **same** direction as the applied emf, thus tending to **maintain** the current. The magnitude of the self-induced or **counter emf** depends upon the time rate of change of the magnetic flux due to the variation of the original current in the coil.

This relation may be given mathematical form. From Eq. (143) it is seen that the magnetic field H is proportional to the current. It has also been shown [Eq. (38)] that the induction B is proportional to H. It therefore follows that the total magnetic flux Φ will be proportional to the current. Hence, if there is no iron in the circuit, we may write

$$\Phi = LI, \tag{180}$$

where L is a constant, the magnitude of which will depend upon the geometrical form of the circuit and upon the number of turns in the coil.

To derive an expression for the self-induced emf we may differentiate the terms given in Eq. (180) with respect to t and get

$$\frac{d\Phi}{dt} = L\frac{dI}{dt}.$$

Since, in general,

$$E = -\frac{d\Phi}{dt},$$

it follows that

$$E = -L\frac{dI}{dt}, \tag{181}$$

where E is the self-induced emf. The negative sign indicates that the relation between emf and current change conforms to Lenz's law.

By rearranging the terms in Eq. (181) one may arrive at a definition of the proportionality constant, thus

$$L = \frac{E}{dI/dt}. \tag{182}$$

From this expression we see that the constant L, known as the **coefficient of self-induction, is the ratio of the self-induced emf to the time rate of change of current.**

A second definition of the coefficient of self-induction may be had from Eq. (180). This would take the form

$$L = \frac{\Phi}{I}, \tag{183}$$

which would warrant the statement that this coefficient is **numerically** equal to the total number of lines of magnetic flux included by the coil when unit current is flowing in the circuit.

From Eq. (182) it follows that the coefficient of self-induction of a circuit has a value of one emu when an emf of one abvolt is developed when the current in the circuit is varying at the rate of one abampere per second. The engineering unit of self-induction is the **henry,** which is equal to 10^9 emu (abhenrys), as defined above. In certain cases the henry proves to be an inconveniently large unit. In such cases the millihenry (one-thousandth part of a henry) is used. The microhenry (millionth of a henry) is also sometimes employed. In view of these relations we may restate our definition of the unit of self-induction as follows: **the self-induction of a circuit has a value of one henry, if the**

counter emf developed is one volt, when the current is changing at the rate of one ampere per second.

The term **self-inductance,** or simply **inductance,** is often used synonymously for the expression **coefficient of self-induction.** It is important to bear in mind that self-induction is a **process** by which an emf is developed, while the term coefficient of self-induction, or inductance, refers to the **condition** that determines the magnitude of the counter emf developed. Self-inductance refers to a certain **property** of the circuit. A self-induced emf **obtains only while the current is changing in value.** A circuit may be inductive regardless of whether it is or is not carrying a current. Further, it should be noted that the relations that we have developed above are based on the assumption that the permeability of the region in which the flux exists is **constant.** If the winding incloses a magnetic material such as iron, for example, whose μ value is not constant, Eqs. (180) to (182), inclusive, do not apply except for very small changes in current.

133. Inductive and Noninductive Circuits. Any circuit that forms a closed loop possesses appreciable inductance. If, however, a short con-

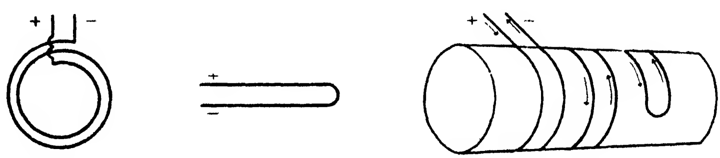


FIG. 179.—Inductive and noninductive windings: (*a*) inductive; (*b*) and (*c*) noninductive.

ductor doubles back on itself without forming a closed loop, it is practically noninductive. Figure 179 illustrates this point. A winding having the form shown in (*a*) will be inductive. In the arrangements represented by (*b*) and (*c*) the magnetic flux due to the current in one wire will be neutralized by the flux due to the return circuit, and hence the circuit, as a whole, possesses practically no inductance. Standard resistance coils are wound in the manner shown in (*c*). If, however, a single straight wire is of considerable length and of large size, it may exhibit a certain amount of inductance.

134. Calculation of Inductance. The calculation of the value of the coefficient of self-induction for any given circuit requires a knowledge of the magnetic flux distribution due to the current in the circuit. Except for a few cases, the analytical treatment is more or less involved and outside the scope of this text. We will, however, outline the treatment for two cases frequently encountered in practice.

We have seen [Sec. 117, Eq. (149)] that the field strength within a long solenoid or helix is given by the expression

$$H = 4\pi n I,$$

where n is the number of turns per centimeter of length of coil. This might be written

$$H = \frac{4\pi NI}{l},$$

where N is the total number of turns and l the length of the solenoid. Using the relations embodied in Eqs. (38) and (41), the **total flux** within the helix would then be given by

$$\Phi = \frac{4\pi NIA\mu}{l},$$

where A is the area of the cross section. Each turn in the solenoid incloses all of this flux; hence the **effective flux** or **flux linkage** that operates to produce self-induction effects will be given by

$$\Phi_{\text{eff}} = \frac{4\pi\mu N^2 IA}{l}.$$

By definition (Sec. 132) the coefficient of self-inductance is numerically equal to the flux due to unit current. Therefore, making I unity, we have the inductance of a solenoid given by

$$L = \frac{4\pi\mu N^2 A}{l}. \tag{184}$$

In many practical cases the cross section of the core is circular; hence

$$L = \frac{4\pi^2\mu r^2 N^2}{l}, \tag{185}$$

where r (the radius) and l (the length) are in centimeters, and L in abhenrys. To express the inductance in **henrys** our relation becomes

$$L = \frac{4\pi^2\mu r^2 N^2}{10^9 l}. \tag{186}$$

In millihenrys, it would take the form

$$L = \frac{4\pi^2\mu r^2 N^2}{10^6 l}. \tag{187}$$

In microhenrys, the equation becomes

$$L = \frac{4\pi^2\mu r^2 N^2}{10^3 l}. \tag{188}$$

If the winding does not include magnetic material, μ becomes unity and we have, for the case of an air-core helix,

$$L = \frac{4\pi^2 r^2 N^2}{10^6 l} \tag{189}$$

in millihenrys.

It should be borne in mind that the permeability μ is a function of the magnetizing field strength, and hence these equations cannot be applied to circuits containing a magnetic substance unless the mean value of μ is known or can be estimated.

It is also to be noted that the foregoing relations were developed on the assumption that the magnetic field within the solenoid is uniform; hence Eqs. (184) to (189), inclusive, hold strictly true only for toroidal windings (Fig. 150). They are, however, applicable to those cases where the length of the coil is great compared with its diameter, say, not less than 10 to 1. Various correction factors have been worked out to take account of the end effects. Professor Nagaoka has developed a correction formula based on the ratio of the radius of the coil to its length and has prepared an accompanying table of constants to assist in making rapid and accurate calculations of the coefficient of self-inductance in those cases frequently met with in practice. Professor Nagaoka's table is to be found in a volume entitled "Calculation of Alternating Current Problems," by Dr. Louis Cohen. These and other correction data are to be found in the *Bureau of Standards Circular* 74, and also in any electrical engineering handbook.

Problem. An air-core coil having 300 turns is 30 cm in length and 2.4 cm in diameter. What is the coefficient of self-induction in abhenrys and in millihenrys?

Solution. Making use of the relation embodied in Eq. (189) we have

$$L = \frac{4\pi^2(1.2)^2 \times 300^2}{30} = 170{,}600 \text{ abhenrys} = 0.17 \text{ mh.}$$

In any given case the total magnetic flux at any point is the algebraic sum of the components due to the several contributing fields. In the case of two parallel wires ("lead" and "return") the total or resultant

field will be the sum of the fields due to each wire. In Fig. 180, let MM' and SS' represent two long parallel wires, as for example, the two leads constituting a telegraph or telephone pair. Let r be the radii of the wires, and d their distance apart between centers. Consider a length l of these wires, and the magnetic field between them. In that part of

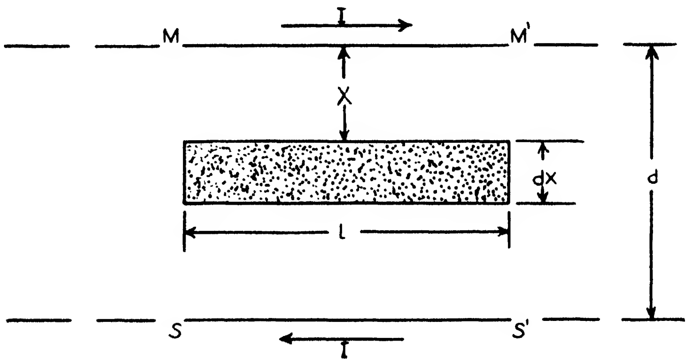


FIG. 180.—Inductance due to two long parallel wires.

the field indicated by the dotted area the total flux is equal to $Hl\,dx$. The total field strength will be equal to [Eq. (141), Sec. 113] the sum of the fields due to each of the two current elements, or

$$\frac{2I}{x} + \frac{2I}{d - x} = H.$$

Recalling our definition of inductance (Sec. 132), we may make I unity, and thus the flux through the shaded area will be

$$\Phi = 2l\left(\frac{1}{x} + \frac{1}{d - x}\right) dx.$$

Considering the whole area between the wires

$$\begin{aligned} \mathrm{L} &= 2l \int_r^{d-r} \left(\frac{1}{x} + \frac{1}{d - x}\right) dx \\ &= 2l\left(\log_e \frac{d - r}{r} - \log_e \frac{r}{d - r}\right) \\ &= 4l \log_e \frac{d - r}{r}, \end{aligned}$$

l and r being in centimeters. Changing to base 10, our equation becomes

$$L = 4l \log_{10} \frac{d - r}{r}\, 2.3026 \quad \text{abhenrys.} \tag{190}$$

In arriving at the above relation we have neglected the field in the wires themselves. If the wires are small compared with their distance apart

the error thus introduced is negligible; in most practical cases this condition obtains.

Problem. A two-wire circuit is 300 m in length, the wires being spaced 30 cm apart. If No. 12 B & S wire is used, what is the inductance of the line?

Solution. Wire of size No. 12 B & S gauge is approximately 0.2 cm in diameter. Utilizing Eq. (190), we have

$$L = 4 \times 30{,}000 \log_{10} \frac{29.9}{0.1} \times 2.3026$$
$$= 7 \times 10^{5} \text{ abhenrys, or } 0.0007 \text{ henry.}$$

Formulas giving the coefficient of self-inductance have been worked out for a number of special cases, particularly in connection with h-f a-c circuits. For a discussion of such cases the student is referred to *Circular of the Bureau of Standards No.* 74, revised edition, or to any standard handbook on communication engineering.

An examination of Eqs. (189) and (190) discloses the fact that the only unit entering into these relations is that of length. It follows, therefore, that the coefficient of self-inductance has the dimension of length, as was the case with capacitance.

In dealing with any quantity such as self-inductance, it is always well to know the magnitude in the case of certain common circuits. Such a knowledge serves to assist one in forming a rough estimate of the values that may be encountered in practice. For instance, the inductance of an ordinary doorbell is something like 0.012 henry; a common telephone receiver (diaphragm in place) varies from 0.075 to 0.1 henry; the secondary of a ¾-in. induction coil measures about 15 henrys; the coil of a sensitive galvanometer is from 1 to 2 henrys; the inductance units frequently encountered in radio communication are of the order of 0.2 mh; the inductance of a mile of telephone line (two No. 10 wires spaced 12 in. apart) is approximately 3.678 mh.

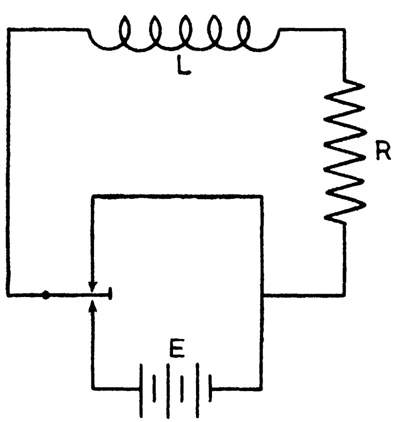


FIG. 181.—Circuit by means of which the effect of inductance on the growth and decay of a current may be studied.

135. Growth of Current in an Inductive Circuit. Reference has already been made (Sec. 132) to the fact that the counter emf of self-induction acts to retard the rise of current in an inductive circuit. It

remains to derive an expression for the value of the current at any time t after an emf is applied to the circuit. In Fig. 181, let E be the impressed emf, R the total resistance and i the value (instantaneous) of the current at any time t after closing the circuit. In order to establish a current in the circuit LR an emf must be applied which shall be equal to the iR drop plus the emf of self-induction. This may be expressed thus,

$$E = iR + L\frac{di}{dt}.$$

A solution of this differential equation will yield an expression giving the current magnitude as a function of time.

To effect a solution of the above equation, we first separate the variables and get

$$\frac{di}{i - (E/R)} = -\frac{R}{L}\,dt. \tag{i}$$

Integrating, we have

$$\log_\epsilon\left(i - \frac{E}{R}\right) = -\frac{R}{L}t + c, \tag{ii}$$

where c is a constant of integration. Equation (ii) may be written in the form

$$i - \frac{E}{R} = \epsilon^{-\frac{Rt}{L}+c}. \tag{iii}$$

At the instant of closing the switch, $t = 0$ and $i = 0$. Hence, from (iii)

$$\epsilon^c = -\frac{E}{R}. \tag{iv}$$

By combining (iii) and (iv) we get

$$i = \frac{E}{R}[1 - \epsilon^{-(R/L)t}] = I_m[1 - \epsilon^{-(R/L)t}], \tag{191}$$

which indicates how the current increases in value with time. I_m represents the maximum value of the current. In the term $\epsilon^{-(R/L)t}$, ϵ is the base (2.7183) of the Napierian system. It is thus evident that the rise of current in a circuit containing inductance and resistance is logarithmic, and will, therefore, require an infinite time to attain a fixed value. It will be seen that when L is small the current reaches a steady value in a very short interval of time.

It has been pointed out (Sec. 134) that inductance has the dimension of length, and it will be shown later that resistance has the dimensions of velocity, viz., length and time. The reciprocal of the ratio R/L would, therefore, involve only the fundamental unit of time. Hence

the ratio $1/(R/L)$ is spoken of as the time constant of a circuit and is commonly represented by the Greek letter τ. Equation (191) may then take the form

$$i = I_m[1 - \epsilon^{-(t/\tau)}]. \tag{192}$$

When $t = \tau$, this reduces to

$$i = I_m - \frac{I_m}{2.7182} = 0.632 I_m. \tag{193}$$

This means that the current reaches 0.632 of its final value in an interval of time equal to the time constant τ. This factor τ is a measure of the growth of the current under the conditions specified. It thus becomes evident that **the rate of increase of current depends not alone upon the inductance L, but upon L and R conjointly.**

136. Decay of Current in an Inductive Circuit. Employing the same designations as in the previous section, let us suppose that, after the current has reached its maximum or steady value, the switch is short-circuited and the battery disconnected. What will be the law of the decay of the current in such a case? Again applying Kirchhoff's law, we may represent the physical conditions by the equation

$$L\frac{di}{dt} + Ri = 0.$$

The solution of this differential equation leads to the relation

$$i = I_m\epsilon^{-(R/L)t}. \tag{194}$$

As in the previous case we may write τ for the time constant term and get

$$i = I_m\epsilon^{-(t/\tau)}. \tag{195}$$

It is therefore evident that the decay of the current in an inductive circuit is also logarithmic; the greater the value of τ, the more slowly does the current diminish in value. Figure 182 depicts the growth and decay of the current in a circuit containing inductance and resistance. It will be noted that, in the particular circuit represented, 0.05 sec was required for the current to reach a steady state and that there was a corresponding lag in the decay of the current after the applied emf was disconnected. In this connection it is interesting to note that the growth and decay of the current are complementary. This is shown by the fact that the sum of $I_m[1 - \epsilon^{-(t/\tau)}]$ and $I_m\epsilon^{-(t/\tau)}$ is I_m, the maximum or steady state of the current.

137. Mutual Inductance. Two or more circuits are frequently associated electrically in such a manner that the magnetic flux produced by one winding threads through the winding of one or more associated

circuits. There is, therefore, a **mutual** magnetic reaction giving rise to an induced emf in one of the coils whenever the current changes in value in the associated circuit. The winding into which energy is being fed is spoken of as the **primary,** and the associated coil, from which

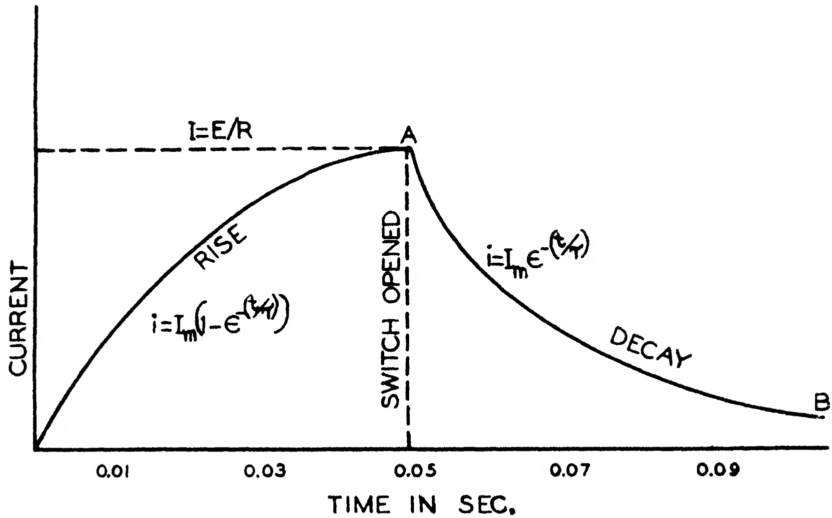


FIG. 182.—Growth and decay of current in an inductive circuit.

energy is being absorbed by some form of electrical load, is referred to as the **secondary.** Such an arrangement of associated circuits has what is called a **coefficient of mutual inductance,** or, more simply, **mutual inductance.** Mutual inductance may be thought of as a **property** of associated circuits by virtue of which any change in the magnitude of a current in one of the circuits gives rise to an induced emf in the other circuit.

The flux in one of the coils, due to a current in the other, will be proportional to that current, or

$$\Phi \propto I.$$

We may write $\Phi = MI$, where M is a constant known as the coefficient of mutual induction. We know that

$$E = -\frac{d\Phi}{dt}.$$

Hence

$$E = -\frac{d(MI)}{dt}$$

$$= -M\frac{dI}{dt}. \tag{196}$$

In words this means that the coefficient of mutual inductance of two circuits is **numerically** equal to the emf developed in one circuit when the current in the other circuit changes at unit rate.

Because of the similarity of the factors involved in self-inductance and mutual inductance, the same unit is used for both quantities, viz., the henry. On this basis the mutual inductance of two magnetically associated circuits will be one henry if an induced emf of one volt is developed in one of the circuits when the current changes at the rate of one ampere per second in the other circuit.

138. Coefficient of Coupling. In general, when two or more windings are magnetically associated, all the flux produced by a current in one of the coils does not pass through the other windings; there is some **flux "leakage."** This is true even when the coils are coaxial and wound on a closed iron core. If the flux linkage is high, the windings are said to be **closely coupled**; if the flux leakage is large, the coils are said to be **loosely coupled.** A so-called **coefficient of coupling** has been set up and is defined by the following relation

$$K = \frac{M}{\sqrt{L_1 L_2}}, \tag{197}$$

where $\sqrt{L_1 L_2}$ represents the geometrical mean of the coefficients of self-inductance of the two windings involved. In the ideal case (zero leakage)

$$M = \sqrt{L_1 L_2}$$

making the coefficient of coupling unity. However, such a condition is never realized in practice. In the case of a transformer (Sec. 146), the coefficient of coupling is an index of the effectiveness of energy transfer from one winding to another. Later, when we come to deal with the air-core transformers used in h-f communication circuits, we shall find that this coefficient is an important factor in the behavior of associated circuits.

139. Measurement of Self-inductance. Since inductance is an important electrical quantity, it becomes necessary to be able to determine its value. There are several methods by which this may be accomplished. The procedure to be followed will depend, somewhat, upon the magnitude of the self-inductance to be measured. If the inductance is of the order of several henrys, and is of such a character that the winding will safely carry several amperes, a voltmeter method (Sec. 161) may be employed. If, however, the inductance is of the order of a fraction of a henry, and is designed to carry currents of low value, some form of bridge network is commonly used. The circuit involved is indicated in

Fig. 183. Such a plan necessitates the use of a standard inductance, a source of alternating current of constant frequency, and a means of detecting such a current. One commonly used form of standard inductance, known as the Brooks inductometer (Fig. 184), consists of two sets

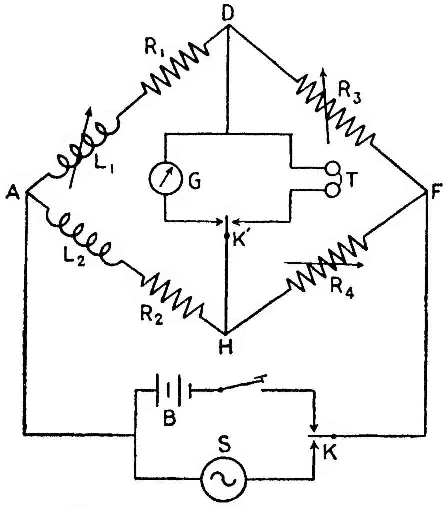


FIG. 183.—Bridge circuit used for measuring self-inductance.

of flat coils, one of which may be moved in a fixed plane with respect to the other.

The method of measurement to be described is based on the fact that, for alternating currents, the loss in potential due to the counter emf is

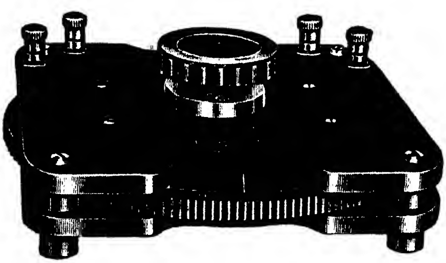


FIG. 184.—Variable form of standard inductance. (*Leeds & Northrup Co.*)

proportional to the magnitude of the inductance over which the drop occurs.

In Fig. 183 let R_1, R_2, R_3, and R_4 be noninductive resistances; L_1 a standard inductance, and L_2 the unknown inductance; S a source of alternating potential difference, usually 1,000 cycles; G a direct current

galvanometer; T a pair of telephone receivers, or other form of a-c detector. The bridge is first balanced for direct current in the usual manner (Sec. 73) by adjusting R_1 and R_2, utilizing the battery B and galvanometer G. Switch K is then thrown to connect the a-c source, and K' closed to connect the a-c detector. If the standard inductance L_1 be of the variable type, its value is adjusted until a balance (silence, or zero deflection) is secured. If the unknown inductance L_2 is of a value which falls within the limits of L_1, R_3 and R_4 may be equal in value. If the unknown inductance is greater or less than the maximum or minimum values of the standard, R_3 and R_4 are adjusted to secure a balance. For this reason the resistances R_3 and R_4 are frequently referred to as the ratio coils. When a final a-c balance obtains, the difference of potential between A and D will equal that between A and H; likewise the drop over DF will equal that over HF. Hence we may write

$$R_1I_1 + L_1\frac{dI_1}{dt} = R_2I_2 + L_2\frac{dI_2}{dt}. \tag{i}$$

The terms $L_1\frac{dI_1}{dt}$ and $L_2\frac{dI_2}{dt}$ represent the loss of potential due to the counter emf developed in the inductances. When a d-c balance obtains,

$$R_3I_1 = R_4I_2. \tag{ii}$$

Differentiating, we get

$$R_3\frac{dI_1}{dt} = R_4\frac{dI_2}{dt}. \tag{iii}$$

Combining (i) and (iii) to eliminate I_2 and dI_2/dt, there results

$$R_4R_2I_1 + R_4L_1\frac{dI_1}{dt} = R_2R_3I_1 + R_3L_2\frac{dI_1}{dt}. \tag{iv}$$

Since the bridge was first balanced for direct current,

$$R_1R_4 = R_2R_3; \tag{v}$$

hence

$$L_1R_4 = L_2R_3, \tag{vi}$$

or

$$\frac{L_1}{L_2} = \frac{R_3}{R_4}. \tag{198}$$

Some form of tuning-fork generator is sometimes employed as the source of a-c potential. The output of a beat-frequency audio oscillator (Sec. 238) is, however, to be preferred. This type of source has the advantage that both the output a-c potential and the frequency of the generator can be readily varied. Under some circumstances it is also

advantageous to replace the headphones by an audio amplifier (Sec. 236), connected to the bridge circuit through a suitable coupling transformer. An additional advantage in using an amplifier, in connection with an inductance bridge, is to be found in the possibility of utilizing an "output" meter instead of a set of telephone receivers for detecting the balance. The personal equation is thereby eliminated. Composite bridges are now commercially available which are designed to measure not only

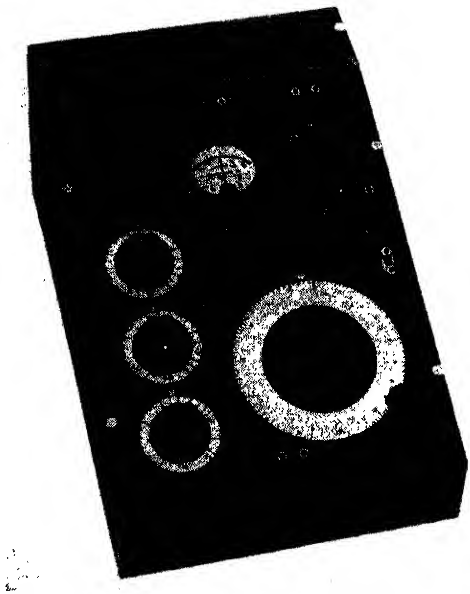


FIG. 185.—Portable form of bridge used for measuring either inductance, capacitance, or resistance. (*General Radio Co.*)

inductance, but capacitance and resistance as well. Such an assembly is illustrated in Fig. 185. This particular unit will measure resistance from 1 microhm to 10,000 ohms; capacitance from 1 $\mu\mu$f to 100 μf; and inductance from 1 mh to 100 henry.

PROBLEMS

1. A coil consisting of 10 turns of wire incloses 10,000 lines of flux. If the flux changes from the maximum value to 2,000 maxwells in 0.1 sec, what average emf will be developed in the winding?

2. A conductor, whose length is 20 cm, moves at right angles to a magnetic field whose flux density is 5,000 lines per square centimeter at a velocity of 1,000 cm/sec. What is the average emf induced in the conductor in emu and in volts?

3. A rectangular coil of wire consists of 10 turns, each of which measures 20 $\times$ 10 cm. The coil revolves at 1800 rmp about an axis parallel to the long

sides, and in a magnetic field where the flux density is 8,000 gausses. Compute the emf for every 15° in one revolution, and plot the values thus obtained against the angular position of the coil.

4. If the coil of Prob. 3 has a resistance of 0.2 ohm, what would be the current in the coil if it were short-circuited when the emf is at a maximum?

5. A certain air-core inductance consists of a closely wound single-layer helix 23 cm in length and 1.7 cm mean diameter. There are 10 turns per centimeter. What is the inductance?

6. A two-wire telephone line is 50 miles long; the wires are No. 10 B & S and are spaced 1 ft apart. What is the inductance of the circuit in henrys?

7. Plot the growth and decay of the current in a circuit having the following constants: Resistance = 0.5 ohm, inductance = 10 henrys, and applied emf = 100 volts.

8. In Prob. 7, what is the value of the current 0.02 sec after the source of potential is disconnected? What is the time constant of the circuit?

9. The core of a certain audio transformer has a volume of 128 cm^3, the thickness of the laminations is 0.6 mm, and the specific resistance of the core material is 10 microhm-cm. If the transformer operates at a frequency of 100 cycles/sec and a flux density of 1,000 gausses, what will be the eddy-current loss? What will be the loss if the frequency is raised to 5,000 cycles?

10. A varying current which changes from 10 amp to zero in $\frac{1}{240}$ sec flows through a circuit which has an inductance of 10 henrys. What will be the value of the self-induced emf?

11. Two windings having inductances of 5 and 10 henrys, respectively, are magnetically coupled. What must be the value of their mutual inductance in order that the coefficient of coupling shall be 0.8?

CHAPTER XVII

GENERATORS

140. Alternating-current Generator. Probably 90 per cent of the electrical energy used is in the form of alternating current. Though the a-c generator, or alternator, as it is commonly called, was not developed until after the d-c machine, it is the more simple unit, being a direct application of the principles outlined in Sec. 130. For these reasons the a-c machine will be considered first.

It is to be kept in mind that a generator is a device for developing or "generating" an emf. By causing a coil to rotate in a magnetic field by mechanical means, an alternating emf is developed in the conductor constituting the winding, and this will, if the coil is suitably connected to an electrical load, cause electrons to move through the complete circuit. Thus mechanical energy may be transformed into electrical energy. In other words, a so-called "generator" is, in fact, a **converter**; it does not, however, produce electrons; instead it provides a means whereby they may be caused to move in definite directions through other conductors. When the rotating coil is not connected to an external load resistance, the potential difference at the terminals of the rotating conductor constitutes the emf of the generator. When the generator winding is connected to a load circuit, the potential difference at its terminals is referred to as the **terminal potential difference** or, more briefly, as the terminal voltage. In general the terminal voltage is less than the open circuit voltage (emf).

In the practical machine, in order to secure maximum magnetic flux in the region of the conductors, the coil is wound on a laminated iron core. The pole pieces are made to conform to the shape of the rotating member, thus producing a field, in the air gap between the poles and the rotor, which is nearly radial and hence at right angles to the moving conductors. The rotating winding and the supporting core is known as the armature. In the rotating-armature type of alternator, the external or load circuit is connected to the rotating coil by means of brushes bearing on so-called slip rings in the manner shown diagrammatically in Fig. 186. These consist of two metallic rings mounted on, but insulated from, the shaft that carries the armature winding. One terminal of the armature is connected to each ring. A photograph of a small armature of this type is shown in Fig. 187. Such a generator would be called a

single-phase alternator. The term "phase" here refers to a single coil, usually made up of a number of turns in series, connected to a single pair of slip rings as shown.

In the case of small alternators the requisite magnetic field may be provided by means of permanent magnets having properly shaped soft-iron pole pieces. (The armature shown in the photograph was part of such an assembly.) However, in machines of any appreciable power rating, electromagnets are provided. This involves an external source of direct current, commonly supplied by a relatively small d-c generator (Sec. 141). The d-c field current is so adjusted that the flux in the magnetic circuit has a value such that the iron will be worked at a point on the B-H curve just below the knee. A rheostat is commonly connected in series with the field winding and the d-c source; this is for the purpose of controlling the field current, and hence the generated emf, in conformity with Eq. (173). The field assembly of a small generator is shown in Fig. 188a. In some generators **two** independent coils, each having their own pair of slip rings, are mounted on the same shaft, the

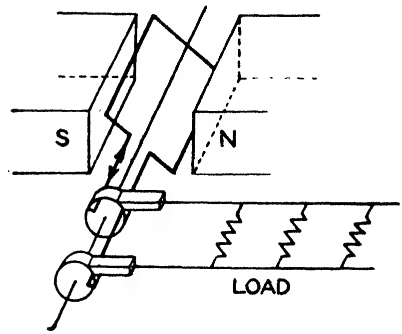


FIG. 186. —Essentials of an a-c generator connected to an electrical load.

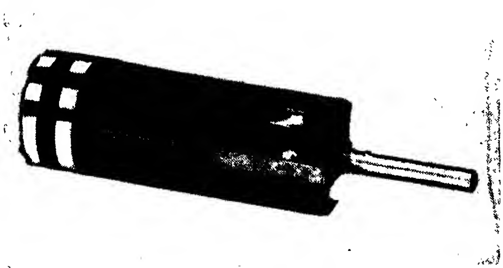
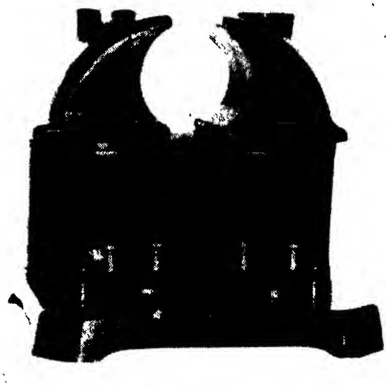


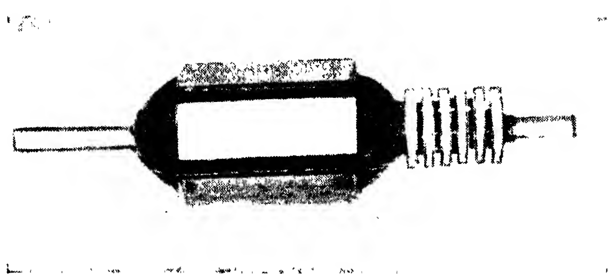
FIG. 187. —Small single-phase armature.

planes of the two coils being placed at an angle of 90° to one another. Obviously, in such a case, the emf generated in one coil will be at a maximum when that induced in the other coil is zero. In other words the two emf will differ in phase by 90°. The phase relation of the emf generated in the two windings is indicated by the curves shown in Fig. 189.

In other alternators **three** independent coils are carried by the same rotor, the windings being positioned 120° with respect to each other.



(a)



(b)

FIG. 188.—Components of a small three-phase a-c generator, rotating armature type.

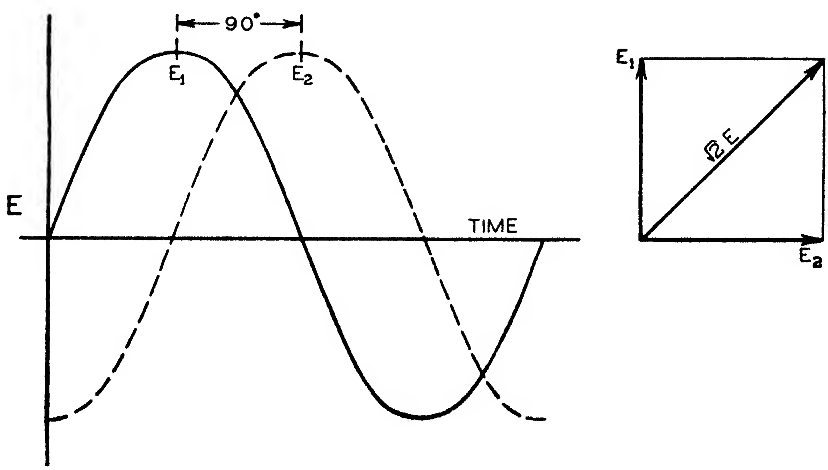


FIG. 189.—Phase relations in the case of a two-phase alternator.

An armature of this type is illustrated in Fig. 188*b*. The phase relations that obtain in such a case are diagrammed in Fig. 190. As indicated above, a machine having one winding is designated as a single-phase unit; one with two coils, a two-phase generator; and a three coil machine, a three-phase alternator. Generators are made that will develop as many as six independent emf. Machines which produce two or more

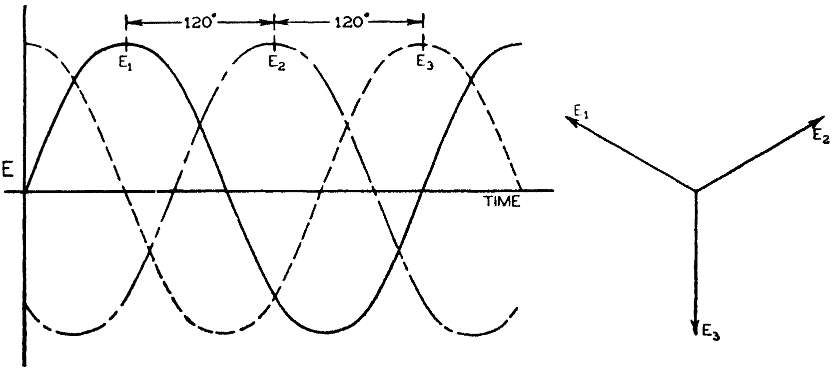


FIG. 190.—Phase relations in the case of a three-phase alternator.

emf are called **polyphase alternators.** Practically all power generation and distribution involves a three-phase, a-c system. The reason for this will be touched upon later.

In generating and distributing three-phase electrical power, it is not necessary to make use of three independent pairs of slip rings and six

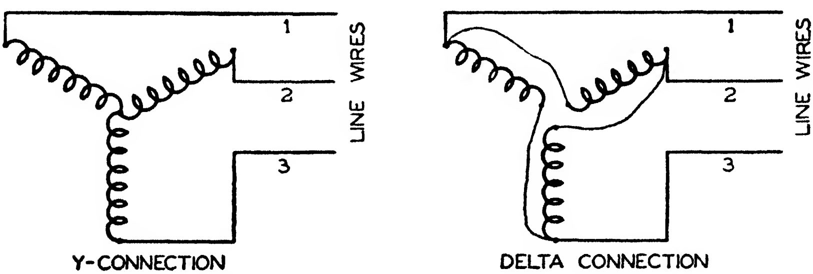


FIG. 191.—*Y* and Δ systems of a-c connections.

separate line wires. Two plans by means of which the several armature windings and the line wires may be arranged to provide a three-wire system are indicated in the two diagrams appearing as Fig. 191. (These two systems of wiring are known as *Y* and Δ connections.) It is to be noted that in such a three-phase, three-wire system of connections, one wire serves as a return conductor for the other two, and that **each** wire,

in turn, functions as a return wire. If we consider instantaneous currents, Kirchhoff's first law holds. It is also to be observed that

$$E_{12} = E_{23} = E_{13}$$

in both the Y and the Δ system. It may be shown that less copper is required to transmit a given amount of power by means of a three-wire system than when using a two-wire network.

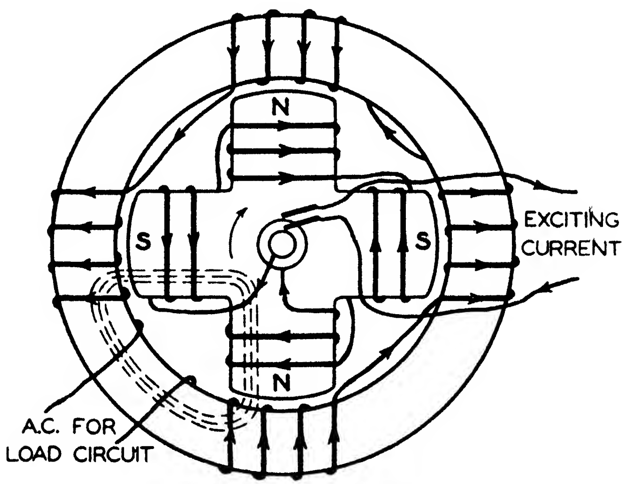


FIG. 192.—Diagramatic sketch of four-pole, single-phase alternator, rotating-field type.

The generators thus far described are of the rotating-armature type. Practically all large alternators, however, are so constructed that the **field rotates** and the armature winding constitutes the stator. Figure 192 is a diagrammatic sketch of a four-pole, single-phase alternator of the rotating-field type. In this type of unit the direct current is led in through slip rings. The outside ring represents the laminated armature frame that, in the actual machine, carries the four sets of armature coils in slots in its inner surface, as shown in Fig. 193. It is seen that the poles alternate and that adjacent armature coils are wound in the reverse direction. As a result, at any given instant, the induced emf in the several coils is additive. If, for instance, the field rotor is revolving clockwise, and has turned through an angle of 90 physical degrees, a pole of opposite sign will have replaced the one shown in the sketch. This means that the induced emf will be reversed in direction. When a rotation of 90° more has taken place, the situation as to polarity will be as originally diagrammed. In short, an electrical cycle has occurred; a change of 360 **electrical** degrees has taken place. It is thus evident that,

in the case cited (a four-pole machine), the emf will alternate four times for every complete physical revolution of the field assembly, *i.e.*, two complete cycles will occur. In general, then, one may determine the frequency of such an alternator (number of cycles per second) by **multiplying the number of revolutions per second by the number of pairs of**

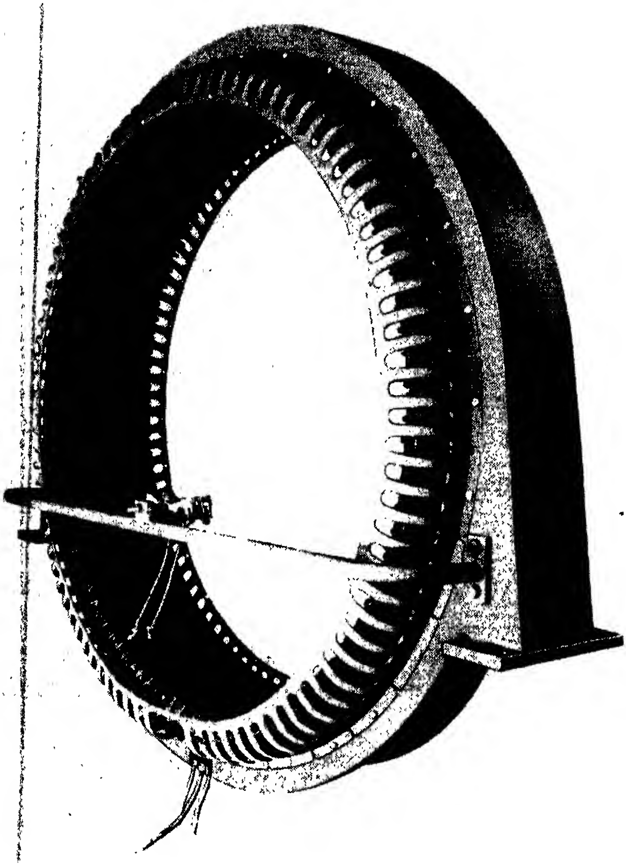


FIG. 193.—Armature (stator) of a polyphase alternator. (*Westinghouse Electric Corp.*)

poles. Figure 194 shows a rotating-field assembly designed to operate with the stator shown in Fig. 193. Figure 195 is an illustration showing a complete rotating-field, three-phase alternator. The small d-c field generator is mounted on the same shaft as the field coils of the alternator. Figure 196 depicts the oscillographic record of the emf developed by one phase of a certain commercial a-c generator.

One advantage of the rotating-field type of generator is that, with

the armature forming the stator element, it is not necessary to pass the load current through moving contacts. Alternators are commonly designed to develop 2,300 volts, or more, in the armature winding. At such voltages moving contact connections are troublesome. The commutation of the low-voltage d-c exciting current presents no particular problem.

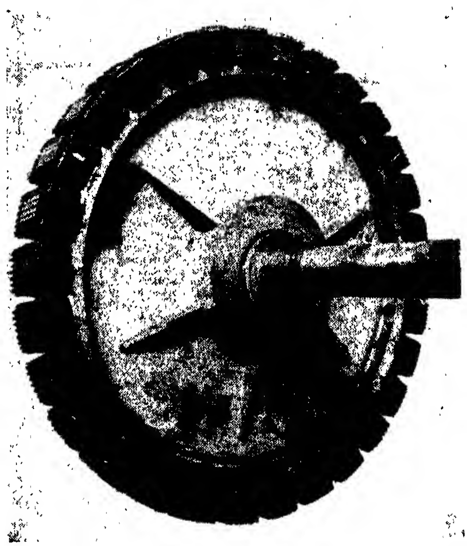


Fig. 194.—Field assembly (rotor) of a polyphase alternator. (*Westinghouse Electric Corp.*)

141. Direct-current Generator. Though by far the greater part of the electrical energy utilized by private and commercial consumers is of the a-c form, there are several limited, but important, fields in which direct current is necessary. For instance, direct current is essential in electroplating, in the electrolytic refining of metals, and in charging automobile storage batteries. Direct-current motors are also better suited to certain types of service than are a-c motive units. Mention has already been made of the use of direct current in connection with the excitation of the field magnets of a-c generators. These and other uses make a generator of direct current necessary.

Fundamentally, the d-c generator does not differ from the a-c unit already studied. In fact, it is essentially an alternator of the rotating-armature type. The emf is inducted in the rotating coil in exactly the same manner in both types of machines. The difference appears in the means whereby contact is made with the rotating circuit. Instead of being continuously in contact with the generator winding, as in the alter-

nator, contact is maintained for only a **part** of each revolution. This is accomplished by means of a device known as a split ring, or **commutator.** This consists of two segments of a single ring, the segments being insu-

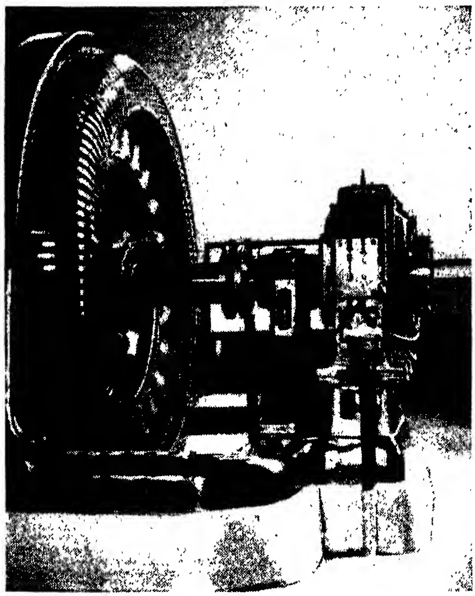


FIG. 195.—Three-phase a-c generator with direct-connected d-c generator for supplying field current.

lated from one another and from the armature shaft. The two ends of the armature coil are connected to these segments as indicated in the diagram of Fig. 197. As the coil rotates, its terminals can be made to

FIG. 196.—Typical wave form of a polyphase alternator.

make contact with a different brush at the instant in the cycle when the induced emf is zero, *i.e.*, as the emf is about to reverse its direction. It therefore follows that when the emf changes direction, the contact with the external circuit also is reversed, with the result that any given brush

is always either positive or negative. In short, rectification is accomplished; the current will always flow in the same direction in the external circuit.

With the single-coil arrangement just described, the emf delivered to the external, or load, circuit would be pulsating in character, as indicated by the full line in Fig. 198. For many purposes a pulsating unidirectional current is not satisfactory. To improve this situation a second coil might be placed on the armature at right angles to the first—as in a two-phase alternator. The terminals of this second coil could be connected to a second set of contact-making segments, as sketched in Fig. 199*a*. The resultant emf would be as shown in Fig. 199*b*. There would still be considerable "commutator ripple," but the emf would never become zero. By providing additional coils, and corresponding commutator bars, the ripple can be reduced still further.

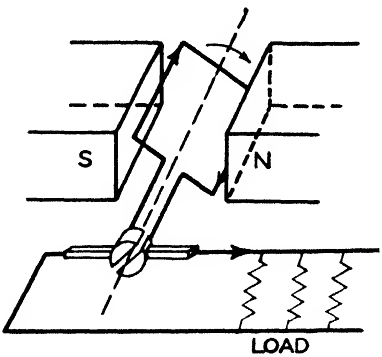


FIG. 197.—Essentials of a d-c generator, connected to an electrical load.

However, with the scheme of armature circuits above mentioned, only one of the coils is "working" at any given time; further, it is contributing to the total emf during only a fraction of the time when it is passing a given pole face, as indicated by the full line curves in Fig. 199*b*. It is possible to arrange an armature winding in such a manner that the above indicated limitations will not obtain. The theory of commutation is a more or less complex subject and beyond the scope of this text. The interested reader will find a full and clear discussion of the subject in "Electrical Engineering" by Dawes, Vol. I, 3d ed., pp. 357 *ff*. Later we shall consider other means whereby alternating current may be rectified.

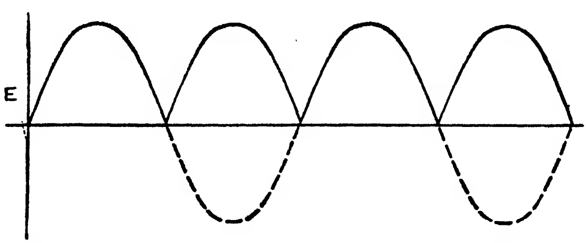


FIG. 198.—Pulsating emf developed by a single-coil d-c generator.

Like the alternator, the field of the d-c machine requires direct cur-

rent for its excitation. This field current may be supplied from a storage battery or a smaller d-c generator, and this is sometimes done. In this event the generator would be spoken of as a separately excited machine. However, in most cases the exciting current is supplied by the machine itself. This may be accomplished in one of three ways as sketched in Fig. 200. In the circuit arrangement shown in Fig. 200*a*, the field winding *F*, consisting of many turns, is connected in parallel, or shunt, across the brushes. Commonly, there is enough residual magnetism in the field

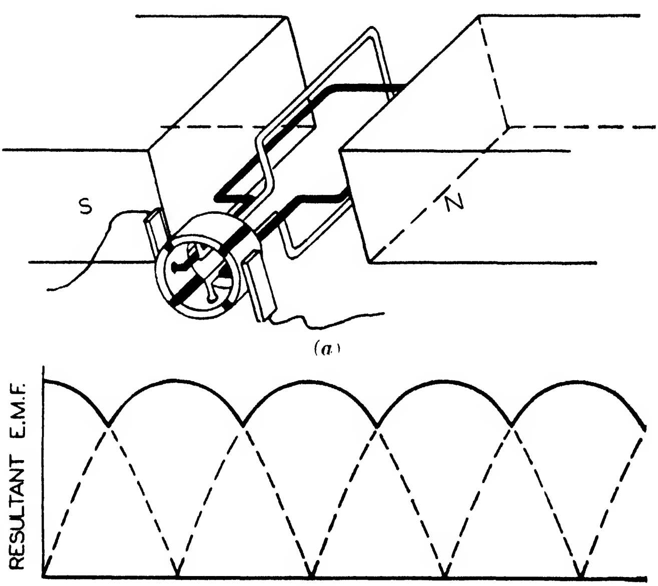


FIG. 199*a* and *b*.—Two-coil d-c generator and resulting wave form. (*After Dawes.*)

magnets to develop a small emf at the brushes when the machine is first started. This emf will produce a corresponding current in the field winding which, in turn, will build up a greater flux in the pole pieces. Unless controlled, this building-up process will continue until the iron in the magnetic circuit is saturated. Usually a variable resistance *R* is inserted in series with the field winding in order to adjust the field current and thus control the terminal voltage of the dynamo. A generator of this type is known as a **shunt** machine. Its load or external characteristic is shown as curve *A* in Fig. 201. The electrical load on a generator is increased by **decreasing** the load resistance. Since the external and field circuits are in parallel, the current through the field winding

will decrease as the load increases, and hence the terminal voltage will fall with increasing load.[1]

A second circuit arrangement for exciting the field is indicated in Fig. 200*b*. In this case the field winding F' is in series with the load

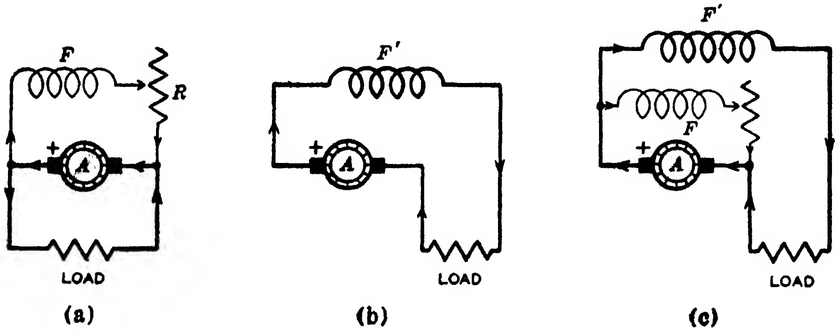


FIG. 200.—Methods of exciting a d-c generator: (*a*) shunt connected, (*b*) series, (*c*) compound.

circuit and the armature winding. All current delivered to the load will, therefore, pass through the field coils. The result is that the field flux will increase in value as the current drawn from the machine increases. The terminal voltage will, accordingly, tend to rise with increasing load,

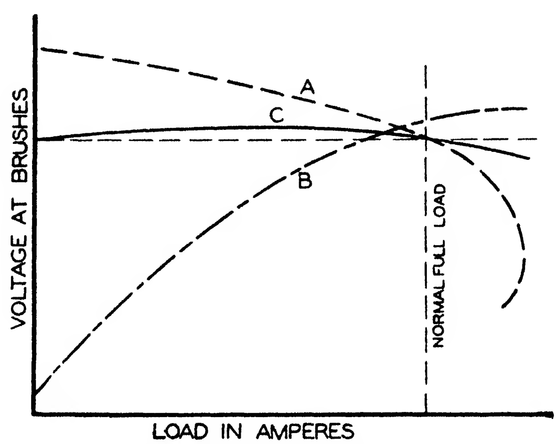


FIG. 201.—Operating characteristics of the types of generators shown in Fig. 200.

as shown in curve *B*, Fig. 201. Such a generator is referred to as a **series** machine, and is said to have a rising characteristic.

[1] There are other factors that operate to cause a decrease in terminal emf with increase in load. For a complete discussion of the performance of d-c generators, the reader should consult any standard text on the elements of electrical engineering.

In some classes of service it is desirable to have a generator whose terminal voltage does not change materially with load. By combining the characteristics of the two previously described machines in a single unit this end can be attained. The field circuits of such a dynamo are sketched in Fig. 200*c*. This type of machine is known as a **compound-wound** generator. The field winding, as the name implies, consists of two parts, one composed of many turns in shunt with the brushes, and a second of a few turns in series with the armature and load. The external characteristic is shown as curve C in Fig. 201, and it will be noted that terminal voltage does not fall appreciably unless the generator is overloaded. The compound machine is the type of generator most commonly used.

In operating any type of generator there is always a drop in potential due to the current in the armature, and the magnitude of this drop will be given by $I(R_a + R_b)$, where R_a is the resistance of the armature winding (usually a fraction of an ohm), and R_b the resistance of the brushes. As the load current increases, the drop will increase correspondingly. Compensation may be effected by adjustment of the field rheostat R.

CHAPTER XVIII

MOTORS

142. Principles of the Direct-current Motor. If a potential difference is applied to the terminals of a d-c generator its armature will rotate; and thus electrical energy may be reconverted into mechanical energy. In other words, it will act as a d-c motor. The fundamental electromagnetic relations involved in such a transformation have been considered in Sec. 118. In that discussion it was pointed out that a rectangular coil through which a current is flowing is acted upon by an electromechanical couple whose value is $IHAN(\cos\theta)/10$ [Eq. (158)], where I is the current (in amperes) through the coil, H the field intensity in the region of the conductors, A the average effective area of the individual turns, N the number of turns in the coil winding, and θ the angle that the plane of the coil makes with the direction of the field. If and when the pole faces conform to the shape of the armature, the magnetic flux is approximately radial, and the above expression for the torque becomes

$$L = \frac{IHAN}{10},$$

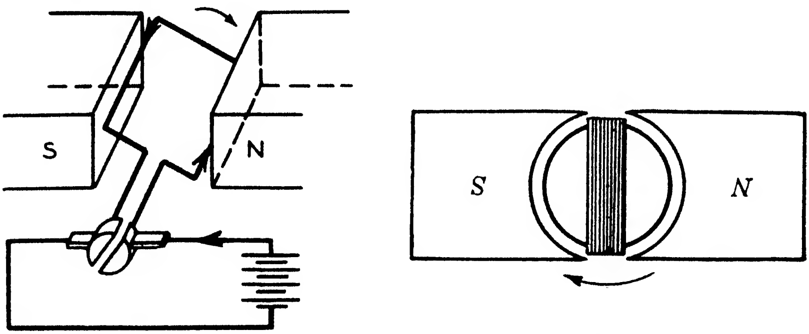
which may be written

$$L = \frac{I\phi N}{10}, \tag{199}$$

where Φ is the total flux threading the coil, I the current in amperes, and L the torque in dyne-cm. The current passing through the armature winding will produce a torque whose magnitude will be given by Eq. (199). By means of a suitable pulley, or clutch arrangement, this torque may be utilized to cause the rotation of a mechanically associated "load."

The electromechanical reactions that take place in a d-c motor may be followed by reference to Fig. 202. Let us assume that a current is sent from an outside source through the armature winding as indicated in (a). In response to the torque developed [Eq. (199)], the coil will rotate in the direction indicated. When the coil reaches the position shown in (b), the torque will be practically zero. Due, however, to the rotational inertia of the armature it will, if not too heavily loaded mechanically, continue to turn past the vertical position. As it does so, the commutator bars shift to opposite brushes, **thus reversing the cur-**

rent through the armature winding. As a result, the torque tending to rotate the armature will still be clockwise, and a continuation of this process will, accordingly, cause the rotor to continue in motion. Thus we have a means whereby a direct current may be converted into mechanical energy, or the converse of what takes place in the case of the generator.



(a) (b)

Fig. 202.—Essentials of d-c motor.

Several aspects of the process just outlined require consideration. As the motor armature coils rotate, they move through a magnetic field, and hence, as in the case of a dynamo, **an emf will be developed in the windings of the rotor.** By Lenz's law the direction of this induced emf will be opposed to the applied emf. Before the armature begins to rotate, the counter emf is zero; hence the current taken by the armature will depend wholly on the magnitude of the applied emf and the value of the resistance of the armature winding. As this resistance is usually only a fraction of an ohm, excessive current would result when the emf is first applied. To avoid this, a "starting resistance" is introduced into the armature circuit until the rotor has attained nearly full speed. As the angular speed increases the counter emf increases, in accordance with the fundamental laws of the generator. The counter emf, however, never equals the applied voltage. If the motor is loaded sufficiently to cause the speed to decrease, the counter emf will correspondingly decrease, with the result that there will be an increase in the armature current. It will thus be seen that a d-c motor is, within limits, self-regulating as to speed. The current taken by the motor will be given by the relation

$$I = \frac{E - E'}{R}, \tag{200}$$

where E is the applied voltage, E' the counter emf, and R the combined armature and brush resistance.

In Sec. 141 reference was made to series, shunt, and compound-wound generators. The fields of constant-current motors are designed in a similar manner, and each type of motor has correspondingly distinct operating characteristics. Series motors develop a large starting torque and are, therefore, used in such service as the operation of electric railways and cranes. Shunt motors operate at nearly constant speed at all normal loads and hence are useful in driving lathes, wood-working machinery, etc. A compound motor gives even more constant speed than the shunt motor. It is used in comparatively small units where constant speed is an important consideration.

143. Revolving Magnetic Field. It will be recalled from one's study of the elements of simple harmonic motion that it is possible to produce circular motion by combining two simultaneous linear simple harmonic motions. The conditions which must be fulfilled in order to attain such an end are (1) that the two component harmonic motions shall be of equal amplitude; (2) that they shall be at right angles to one another; and (3) that they differ in phase by 90°.

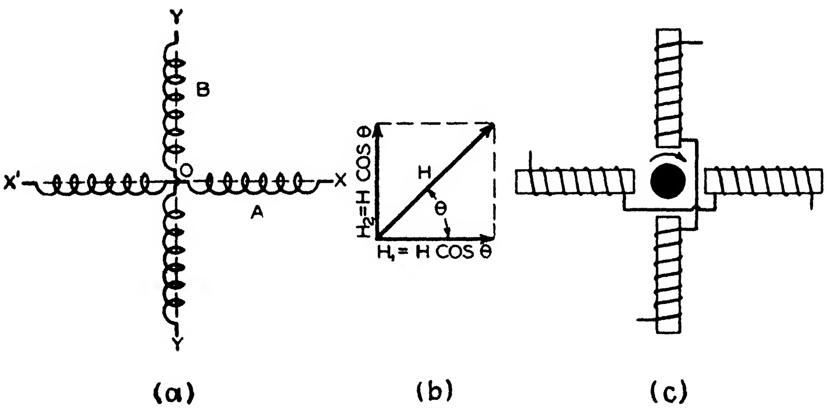


FIG. 203.—Production of a rotating magnetic field.

Since alternating currents, and the resulting magnetic fields, obey the laws of simple harmonic motion, it is possible to produce a rotating magnetic field by combining two alternating fields.

Suppose two coils A and B, carrying alternating current of the same frequency and amplitude, are positioned at right angles to one another as shown in Fig. 203a. Assume that the current in coil B differs in phase by 90° from the current in coil A. The alternating current in A will give rise to a harmonic field H_1 along XX' and the current in B will produce a corresponding sinusoidal field H_2 along YY'. Since the currents differ in phase by 90°, these two fields will also be in quadrature.

The character and magnitude of the field that results from two such component fields may be determined by finding the resultant H according to the usual vector methods. Referring to Fig. 203b, the instantaneous value of the field H in the direction OX, due to the current in coil A, will be given by

$$H_1 = H_m \cos \theta,$$

where H_m is the maximum value of the field in that direction. The corresponding value of the field H_2 along OY will be expressed by

$$H_2 = H_m \sin \theta.$$

The resultant field will then be

$$H = \sqrt{H_1^2 + H_2^2}.$$

By substitution we get

$$H = H_m \sqrt{\cos^2 \theta + \sin^2 \theta}.$$

Since, in general, $\cos^2 \theta + \sin^2 \theta = 1$, it follows that $H = H_m$. It is evident, therefore, that the **resultant field is constant in magnitude.** Further, the instantaneous position of the vector representing this resultant field H will be given by the relation

$$\tan \theta = \frac{H_2}{H_1} = \frac{H_m \sin \theta}{H_m \cos \theta} = \tan \theta = \tan (\omega t).$$

Accordingly it is evident that **the resultant field H rotates with an angular speed of ω radians per second**; and since $\omega = 2\pi f$, the resultant field will rotate a number of times per second corresponding to the frequency of the current in the coils A and B. In other words, the revolving field will rotate in **synchronism** with the alternator that is supplying the windings A and B. Thus we see that a two-phase alternating current can be made to develop a revolving magnetic field. By an extension of the above reasoning, it may be shown that one may produce a rotating field by means of a three-phase current.

144. The Induction Motor. In our discussion of eddy currents (Sec. 131), it was pointed out that a conductor of considerable area, when in the presence of a moving magnetic field, will tend to undergo physical displacement due to the Foucault currents established in the body of the conductor.

We have just seen that a rotating field of fixed magnitude and constant angular velocity may be established by combining the fields due to two or more alternating currents. Assume that a rotor made up of conductors of low resistance is mounted on a shaft in such a manner

that it will be free to rotate within the region of a rotating field produced as outlined above (Fig. 203c). The field due to the eddy currents generated in the rotor by the rotating field will, in conformity with Lenz's law, cause the rotor to revolve and thus develop motor action. This is the basis of the **induction motor.** Nikola Tesla was the first to apply these principles in the production of a practical motor of this type. In practice the rotor of the induction motor commonly consists of a series of heavy copper bars supported on the periphery of an iron framework and parallel to the axis of rotation. The conducting bars are short-circuited by being welded to a copper or brass ring at each end, thus forming what is frequently referred to as a "squirrel-cage" rotor (Fig. 204). The field coils are usually supplied with either two- or three-phase alternating current. In the three-phase units, three sets of field

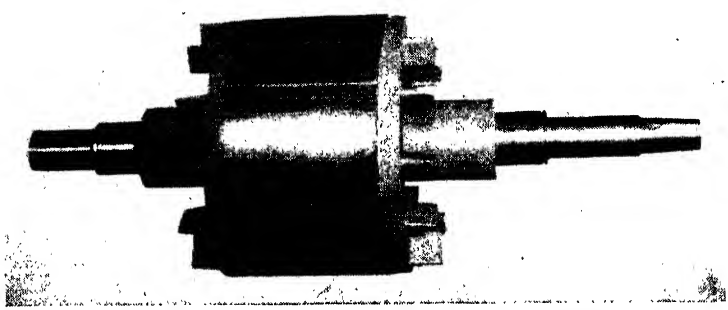


FIG. 204.—Rotor of an induction motor. *(General Electric Co.)*

coils are so arranged on the stator that they produce three fields spaced 120° apart, and also differing in phase by 120°. Such motors are extremely rugged in construction, and because of the absence of a commutator or a slip ring, require little operating attention. Two- and three-phase induction motors are widely used, especially the latter type, the units ranging in size from a fraction of 1 hp to 200 hp.

If a polyphase induction motor is running at normal speed, it will continue to run and carry a mechanical load even though all but one of the phases are disconnected. Under such circumstances the unit is operating as a **single-phase** motor. Such a motor is, however, not inherently self-starting; the rotor must be brought up to near-synchronous speed before the motor will function. Any one of several means may be provided to give the rotor this initial angular velocity. One plan amounts to the splitting of a single-phase current into two components which differ in phase. This can be done, for instance, by providing two windings on the stator, one of which is supplied with an inductance or capacitance in series with that particular phase winding.

The presence of the reactance (Sec. 162) will cause the current in that particular winding to lag, or lead, the current in the other winding, and thus to differ in phase, with the result that a rotating field is established. Figure 205 illustrates the fundamentals of this particular starting scheme. When the rotor attains operating speed, a centrifugal device opens the auxiliary or starting winding F'.

It is evident that, if such a motor will run and carry a load once the rotor is brought up to speed, a revolving field must exist. It would take

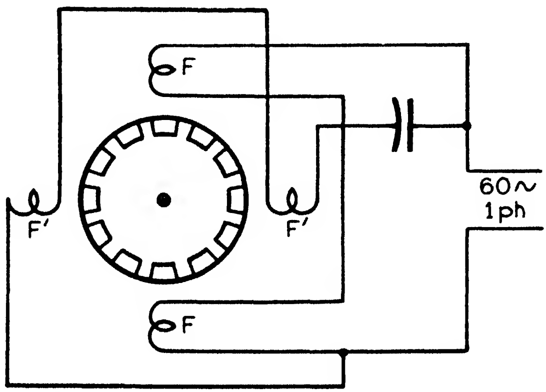


FIG. 205.—Essentials of a single-phase induction motor.

us too far afield to discuss the theory involved in this action, but the interested reader will find a detailed and clear treatment of this subject in a work entitled "Elements of Electrical Engineering" by A. L. Cook, 3d ed., Chap. XXXIV. Other references that will be helpful are to be found in "Electrical Engineering" by C. L. Dawes, Vol. II, Chap. X, and in "Electrical Engineering" by C. V. Christie, p. 482.

Single-phase induction motors resemble the polyphase induction motor in their structural features. The single-phase type of unit is usually made in fractional horsepower sizes only. They are extensively used for domestic purposes, since they will operate on the usual single-phase lighting circuit.

145. The Synchronous Motor. If alternating current is supplied to the armature winding of a polyphase generator whose field is excited from a d-c source and whose rotor is brought up to speed, it will continue to run and will carry a mechanical load. In other words, such a machine is a dynamo in the strict sense of the term; it will function either as a generator or as a motor. It will, however, only function as a motor when running at synchronous speed; hence the name.

When a polyphase voltage is applied to the armature windings of a synchronous motor, a revolving magnetic field will be developed, as in the

case of the induction motor described in the preceding section. If, then, the rotor carrying the field magnets is brought up to synchronous speed and the field current applied, the rotor will lock into step with the revolving magnetic field developed by the armature winding. If overloaded mechanically, it will fall out of synchronism and stall. The synchronous motor is, therefore, a **constant speed** unit; hence it is not suited for use where the load is variable in magnitude. They are extensively used, however, particularly in large units.

CHAPTER XIX

THE TRANSFORMER

146. Fundamental Theory of the Static Transformer. Perhaps the most widely used device in the field of applied electricity is the static transformer. The term "static" signifies that there are no moving parts, potential transformations being effected by changes in magnetic flux within fixed windings grouped about a common iron core. The possibility of the commercial distribution of electric power over extended areas is largely due to this agency. In communication engineering, particularly in telephone and radio practice, the transformer also finds wide application.

In its simplest form a transformer consists of a closed magnetic core upon which are wound two coils (Fig. 206), one usually having a greater

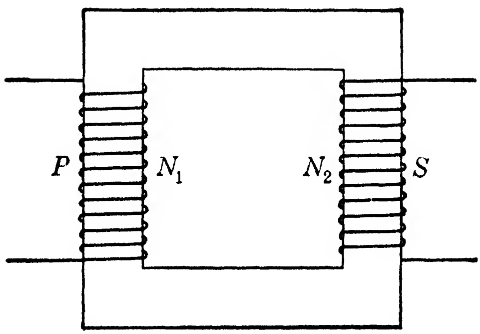


Fig. 206.—Static transformer.

number of turns than the other. The winding **to which** energy is supplied is known as the **primary** and the coil **from which** energy is taken is known as the **secondary.** It is important that this distinction be kept clearly in mind, as the transformer may be utilized to increase ("step up") or to decrease ("step down") the potential. It may be, and often is, used simply as a coupling agent between two parts of an electrical organization, in which case no voltage transformation occurs.

We will first examine the relation which obtains between the potential differences at the terminals of the two windings. Let us assume that sufficient current is flowing in the primary, as the result of an applied alternating emf, completely to magnetize the iron core, and that the

secondary winding is open. As in dealing with a simple inductive circuit, the applied emf must equal the sum of the ohmic drop in the primary winding (in this case due to the magnetizing current) and the counter emf of self-induction. This may be expressed thus,

$$e_1 = Ri_1 + L_1\frac{di_1}{dt}, \tag{i}$$

where the subscripts indicate conditions in the primary winding.

It has been previously shown (Sec. 132) that the emf of self-inductance is equal to the rate of change of the flux. Hence we may write

$$L_1\frac{di_1}{dt} = \frac{d\Phi_1}{dt}, \tag{ii}$$

where Φ_1 = the total flux existing in the iron due to the current in the primary. We are interested in the relation between the number of turns and the existing flux. We have the relation

$$\Phi_1 = N_1\Phi, \tag{iii}$$

where Φ indicates the flux due to each turn and N_1 the turns in the primary winding, the term $N_1\Phi$ being what is called the **"flux turns"** of the primary. Differentiating (iii), we have

$$\frac{d\Phi_1}{dt} = N_1\frac{d\Phi}{dt}. \tag{iv}$$

In view of the relations expressed by (ii) and (iv), we may put (*i*) in the form

$$e_1 = R_1i_1 + N_1\frac{d\Phi}{dt}. \tag{v}$$

Since in a transformer that part of the total applied emf which is required to supply the Ri drop is very small compared with that necessary to balance the reactance drop (Sec. 161) (in practice about 1 per cent), we may neglect the Ri term, and, for our purposes, consider that

$$e_1 = N_1\frac{d\Phi}{dt}. \tag{vi}$$

In general,

$$\text{Emf} = \frac{d\Phi'}{dt} = e',$$

or

$$d\Phi' = e'\,dt;$$

hence

$$\Phi' = \int e'\,dt; \tag{vii}$$

the primes signifying values in general.
From (iii) we have

$$\Phi = \frac{\Phi_1}{N_1}, \tag{viii}$$

which gives the flux linkage per turn.
Combining (vii) and (viii), we have

$$\Phi = \frac{1}{N_1}\int e_1\,dt. \tag{ix}$$

To evaluate the expression $\int e_1\,dt$, we will assume that the applied emf is a sine function, thus

$$e_1 = E_1 \sin \omega t.$$

Substituting this value for e_1 in (ix),

$$\Phi = \frac{E_1}{N_1}\int \sin \omega t\,dt,$$

which yields

$$\Phi = \frac{-E_1}{N_1\omega}\cos \omega t. \tag{x}$$

We thus have an expression for the **flux per turn** in terms of the maximum applied emf and the number of turns in the primary.

Passing now to a consideration of what takes place in the secondary, it may be noted that the **flux turns** of the secondary may be written

$$\Phi_2 = N_2\Phi,$$

and hence

$$\frac{d\Phi_2}{dt} = N_2\frac{d\Phi}{dt}. \tag{xi}$$

But

$$e_2 = \frac{d\Phi_2}{dt},$$

which from (xi) becomes

$$e_2 = N_2\frac{d\Phi}{dt}. \tag{xii}$$

We may eliminate $d\Phi/dt$ by differentiating (x) and substituting in (xii). This yields

$$e_2 = \frac{N_2}{N_1}E_1 \sin \omega t. \tag{xiii}$$

But

$$E_1 \sin \omega t = e_1;$$

hence (xiii) becomes

$$e_2 = \frac{N_2}{N_1} e_1,$$

or

$$\frac{e_1}{e_2} = \frac{N_1}{N_2}. \qquad (201)$$

It is thus evident that the **ratio of the applied and developed potential differences is equal to the ratio between the turns in the primary and secondary.** This, it should be noted, is on the assumption that **all** of the flux produced by the current in the primary is linked with the secondary. In a well-designed transformer the flux leakage is very small; hence Eq. (201) serves as a useful working relation.

Thus far we have not considered the **effect** of the current which may exist in the **secondary** if it is connected to a load impedance. When current flows in the secondary, the magnetic effect of this current, by Lenz's law, will oppose the flux in the core due to the magnetizing current in the primary. This means that the counter emf due to the self-inductance of the primary is lessened, with the result that a greater current will tend to flow in the primary winding. A transformer thus becomes largely self-regulating, the flux not decreasing more than 1 per cent between no load and full load. It will thus be evident that if the load impedance of the secondary is made zero ("short-circuited"), not only will the secondary winding be destroyed, but, due to excessive current in the primary winding, it also may be damaged.

The magnitude of the potential difference developed at the terminals of the secondary of a transformer may be expressed in terms of the flux, the frequency, and the number of secondary turns. The harmonic flux in the core due to the magnetizing current in the primary threads through the secondary winding and induces therein an emf whose value is given by

$$e_2 = -N_2 \frac{d\Phi}{dt}.$$

Assuming that the flux is sinusoidal and has a maximum value Φ, we may write

$$\begin{aligned} e_2 &= -N_2 \frac{d}{dt} (\Phi \sin \omega t) \\ &= -\omega N_2 \Phi \cos \omega t; \end{aligned}$$

which may be written

$$e_2 = 2\pi f N_2 \Phi \sin (\omega t - 90).$$

Thus it is seen that the induced emf is also sinusoidal and lags the flux by 90°. The maximum value will be

$$E_m = 2\pi f N_2 \Phi,$$

and the effective value (Sec. 156)

$$E_2 = \frac{2}{\sqrt{2}} \pi f N_2 \Phi = 4.44 f N_2 \Phi,$$

where E_2 is in abvolts. In engineering units

$$E_2 = 4.44 f N_2 \Phi 10^{-8} \text{ volts,} \tag{202}$$

which reduces to

$$E_2 = 4.44 f N_2 A B 10^{-8} \text{ volts,} \tag{203}$$

where A is the cross-sectional area of the core and B the flux density. It may be shown that

$$E_1 = 4.44 f N_1 A B 10^{-8} \text{ volts.} \tag{204}$$

Equations (203) and (204) are basic relations in transformer design.

Modern power transformers have an efficiency of the order of 95 per cent, so that, approximately,

$$P_1 = P_2,$$

and, except for small copper and iron losses, one might write

$$e_1 i_1 = e_2 i_2,$$

or

$$\frac{e_1}{e_2} = \frac{i_2}{i_1}. \tag{205}$$

The wide utility of the transformer in electrical power distribution rests upon the relations given as Eqs. (201) and (205). Electrical energy may be transmitted at high potentials and low current values and "stepped down" for the consumer's use. In power work this procedure results in great savings in the cost of line conductors, though the cost of insulation is considerably increased. The net gain, however, is large. This economy results from the fact that thermal losses in the line vary as the square of the current.

The phase relations in the transformer deserve consideration. These relations, for a loaded unit, are set forth in the simplified vector diagram appearing as Fig. 207. E_1 represents the emf applied to the primary winding, $I\phi$ the magnetizing current in the primary (which lags the applied emf by nearly 90°), and ϕ the magnetic flux threading both windings. I_1' is that component of the primary current which is required

to neutralize the demagnetizing effect of the load current I_2 in the secondary and I_1 is the total primary current of the loaded transformer. It is to be noted that the emf E_2 developed in the secondary is opposite in phase to the applied emf. It is important that this fact be kept in mind, particularly when dealing with communication circuits. The phase angle α_2 which the load current in the secondary makes with the terminal potential difference of the secondary will depend for its value on the constants of the load circuit; further, the phase angles α_1 and α_2 are, in general, not equal.

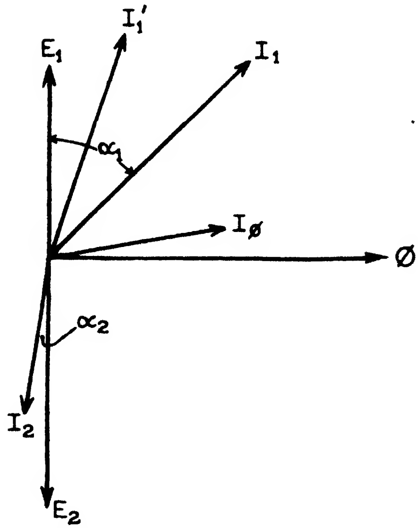


FIG. 207. -Phase relations in a transformer.

147. Types of Transformers. Transformers are of three general designs (Fig. 208) which are commonly designated as the **core type,** the **shell type,** and the **open-core type.** In the former (*a*) the core consists of a single continuous magnetic path (rectangular or circular in cross section), the primary and secondary being disposed about different sections of the core. In certain forms of this type both the primary and secondary are divided into

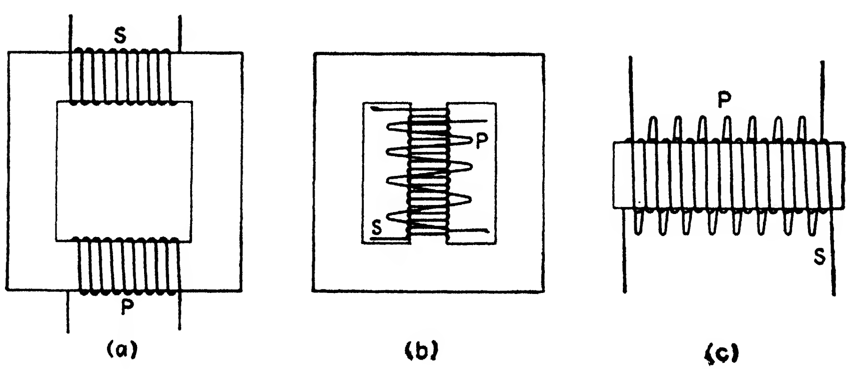


FIG. 208. Types of transformers: (*a*) core type, (*b*) shell type, (*c*) open-core type.

two sections; a primary and a secondary component being placed on separate sides of a rectangular core. In other forms, both the primary and secondary are concentrically wound on the same "leg," or all the

primary may be about one part of the core and the entire secondary about another section of the iron.

Transformers of the shell type (*b*) differ from the core type in that windings are disposed about a central section of the iron, and the magnetic circuit is completed through two or four paths. In some cases the primary and secondary are concentric, while in others both the primary and secondary are arranged in alternate sections, or "pied," as it is sometimes called. Power transformers are constructed in both the core and shell types.

The open transformer (*c*) is the most simple of all the types, consisting of a straight core upon which the primary and secondary are concentrically disposed. This type is usually made only in small sizes, and is used primarily for certain types of telephone equipment.

There are various modifications of the above-mentioned forms. The particular type of transformer that is to be employed in a given case depends largely upon the special use to which it is to be put.

148. The Autotransformer. A special type of transformer which has recently come into wide use consists of a **single** winding disposed about an open or a closed core. The principle involved in the operation of the autotransformer is shown diagrammatically in the sketch appearing as Fig. 209. As commonly designed and employed, the input connections are so arranged that all, or nearly all, of the winding is connected to the source. One output connection S_1 is common with one of the input leads P_1. The second output connection S_2 is variable and is so arranged that any fraction of the total drop in potential over the entire winding may be included between S_1 and S_2. The drop between S_1 and S_2 is due largely to the inductive reactance (Sec. 161) involved. These units are commonly employed as step-down transformers, but gain in voltage may be had by including **more** turns between the output taps than are included in the input connections. Since the currents in the primary and secondary circuits are of opposite phase, that part of the winding that is common to both circuits carries a current whose value is the **difference** (vector sum) between primary and secondary currents. It is evident that, in this type

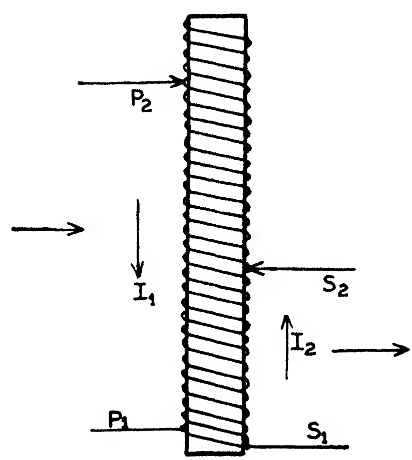


FIG. 209.—Circuits of the autotransformer.

of transformer, the whole of the energy does not undergo transformation. For instance, in the case of a step-down unit of this type, the transformer serves to supply the required increase in current by subtracting from the applied emf. As a result of this relation, the autotransformer is more efficient than a unit of the standard type, and the gain in efficiency becomes more evident as the ratio of transformation approaches unity. In this type of transformer

$$\frac{E_1}{E_2} = \frac{N_1}{N_2} = \frac{I_2}{I_1}. \tag{206}$$

In practice, the winding is usually toroidal in form. By means of a wiping contact, controlled by a central shaft, contact may be made with each turn of the winding. Thus extremely small changes in output voltage may be secured. A typical unit of this type is illustrated in Fig. 210. Such transformers are employed in connection with rectifying outfits (Sec. 232), in balancing units in certain power-distributing systems, and for the control of theater lighting instead of cumbersome and troublesome rheostats. Autotransformers find wide use in laboratory and research work, as well as in industrial processes. When using this type of voltage control, only a trifling amount of energy is lost in the form of heat, and the voltage can be adjusted in imperceptible steps. Autotransformers available for commercial use in various capacities ranging from a few watts to several kilowatts.

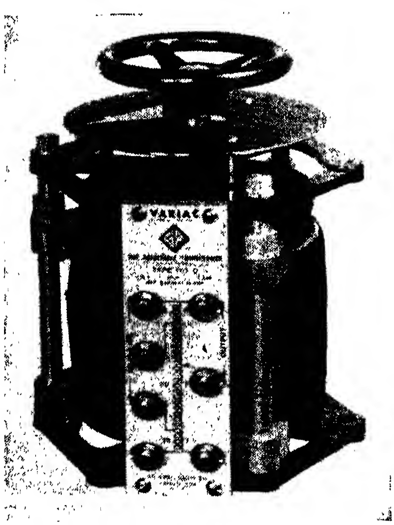


Fig. 210.—Commercial form of autotransformer) without case. (*General Radio Co.*)

149. Transformer Efficiency and Regulation. The efficiency of the ordinary power transformer is very high at normal load, varying from 90 per cent in the smaller units to 98 per cent in the larger sizes. In general, the efficiency of a transformer, expressed in per cent, may be defined as

$$\frac{\text{Output}}{\text{Input}} \times 100 = \frac{\text{output}}{\text{output} + \text{losses}} \times 100.$$

In applying the above relation, it is necessary to specify whether instantaneous or all-day efficiency is meant. By all-day efficiency of a trans-

former is meant the ratio of the total energy delivered by the transformer during a 24-hr period to the total energy delivered to the unit during the same period of time, usually expressed in per cent. The energy input and output would ordinarily be expressed in kilowatt-hours. The reason for differentiating between the efficiency at any moment and the all-day rating is that the primary of a power transformer in a distributing system is connected to the mains at all times, whether it is carrying a load or not; hence the iron losses are continuous.

In determining the efficiency of a transformer, it is common practice to determine the losses, rather than to measure the total output and input. This procedure is followed because one-half of 1 per cent error in a wattmeter reading would cause an error of several per cent in the results. This results from the fact that the efficiency, in most cases, is very high.

By **regulation of a transformer** is meant the relation of the secondary voltage at full load to that at no load, when the input voltage is held constant. Expressed in terms of percentage, regulation may be defined as

$$\frac{\text{No-load voltage} - \text{Full-load voltage}}{\text{Full-load voltage}} \times 100.$$

In cases of small units the regulation is of the order of 2.5 per cent and for the larger sizes about 1 per cent. The regulation depends chiefly on the resistance drop.

150. Constant-voltage Transformers. Another type of special transformer has recently been introduced that is designed to deliver a constant-

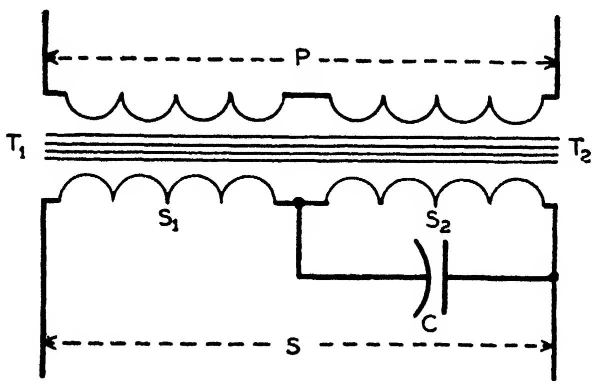


FIG. 211.—Circuits of constant-voltage transformer.

voltage output even though the input voltage may vary between quite wide limits. This automatic regulation is brought about by means of a multiple winding in both the primary and secondary circuits. The

magnetic reactions are somewhat complicated and outside the scope of this volume. However, a general understanding of the principles involved in such a unit may be had by reference to Fig. 211. Such a stabilizing organization, in effect, consists of two transformers, the two primaries and the two secondaries being in series as shown. One of these transformers T_2 operates at high magnetic density. Its secondary is resonated (Sec. 164) by means of the condenser C. The secondary windings are opposed, one winding S_1 being considered as a compensating element to the other S_2. The compensating secondary S_1 has few turns compared with S_2. If the primary voltage increases, for instance, the output voltage in both secondary windings will increase, but since the voltage developed in these two windings differs in phase by 180°, the resultant voltage will tend to remain constant. Alone, this compensation would not result in a strictly constant output voltage. However, the combined effect of controlled flux density and resonance gives a vector sum which is very nearly constant. Commercial units have recently become available that will hold the output voltage within $\pm\frac{1}{2}$ per cent when the applied emf varies from 95 to 130 volts; change in load values does not modify these limits. It is said that the automatic adjustment takes place within a fraction of a cycle. Because of the high flux density in the core, and because of the resonance effect, the output wave form is somewhat distorted. For certain uses this characteristic would constitute a serious limitation, but, for many purposes, a slight increase in harmonic content would be of no consequence. Constant-voltage transformers of this type are finding use as voltage-stabilizing devices. They are made in units ranging in size from a few watts to several kilowatts.

151. Constant-current Transformer. Thus far in our discussion, we have dealt only with transformers designed to give a constant potential. There is, however, a class of service, particularly in illuminating engineering practice, in which it is desirable to have available a transformer unit which will produce a **constant current.** Such a transformer is shown diagrammatically in Fig. 212. It will be noted from the drawing that the secondary is movable and so suspended that it is nearly balanced by a counterweight. When the transformer is in operation, the currents in the primary and secondary, at any given instant, are in opposite direction. Hence, the magnetic field will cause repulsion between the coils, and the secondary will be forced upward and away from the primary. This gives rise to a greater magnetic leakage between the two coils with a resulting decrease in the voltage developed in the secondary. This tends to decrease the current in the secondary and the connected load circuit, thus lessening the magnetic repulsion. The falling of the

secondary will then immediately readjust the voltage conditions so that a constant current is automatically maintained. The constant-current transformer is used in supplying energy to series street lamps.

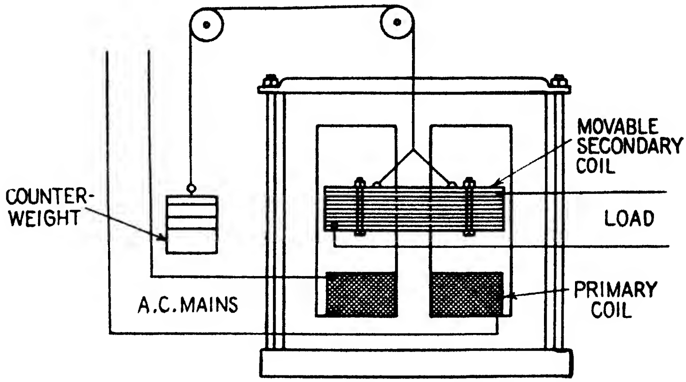


FIG. 212.—Essentials of constant-current transformer.

152. Instrument Transformers. When measuring potential differences in connection with high-potential a-c circuits, it is the practice to connect the voltmeter to the line through a potential or instrument transformer as shown in Fig. 213*a*. It is customary to step down the potential to 110 volts. The output of such a transformer is only sufficient to operate one or two voltmeters.

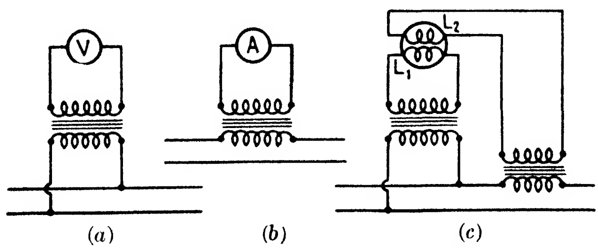


FIG. 213.—Methods of connecting instrument transformers.

In measuring current in high-tension a-c circuits, it is impracticable to employ shunts; therefore recourse is had to what is known as a **current** (or series) **transformer.** The manner of connecting such a device is shown in Fig. 213*b*. In dealing with current values of the order of 1,000 amp, the primary consists simply of the main conductor itself. The winding is so arranged that the secondary will deliver a few watts and is figured on the basis of 5 amp in the secondary.

In the case of wattmeters and power-factor meters for use on high-voltage circuits, a combination of two such instrument transformers is

employed. A sketch of the connections for a wattmeter layout is shown in Fig. 213c. The potential coil of the wattmeter is designated L_1 and the current winding as L_2.

153. Welding and Furnace Transformers. In the operation of electric furnaces and in electric-welding processes, current values as high as 25,000 amp may obtain. In order to secure currents of this magnitude, use is made of special step-down transformers having a few turns on the primary side and a single heavy-copper conductor for the secondary or load circuit.

154. Audio Transformers. Communication engineering covers a wide and rapidly expanding field. The varied aspects of telephone and telegraph transmission, of radiotelegraphy and radiotelephony, and of television involve vast numbers of relatively small transformers, most of which operate at audio frequencies. There are, literally, millions of such units in daily use. It is to be noted that the theoretical problems involved in the design of transformers used for communication purposes are far more complex than those encountered in power generation and distribution. The reason for this is that in power work a given transformer is designed to operate at a definite and fixed frequency and wave form, while in the case of communication transformers the unit must function efficiently over a relatively wide range of frequencies when supplied by a current having a complicated and varying form. Since the fidelity of speech and musical reproduction depends upon the preservation of the current wave form, it is evident that the problems involved are of an entirely different order than those encountered in power practice. While it is outside the scope of this text to enter into a detailed discussion of the many special types of transformers employed in the communication field, mention may be made of two or three of the forms most commonly encountered in practice.

The simple open-core transformer, used in many desk and wall telephone sets, has already been referred to. This unit serves to transform the variable direct current, produced by the action of the sound waves upon the microphone button, into an alternating current of higher voltage. The device is commonly spoken of as an induction coil, rather than as a transformer, because the primary, as just noted, is supplied by a direct current (variable) rather than by an alternating current. Fundamentally, however, it is a transformer, and the principles of transformer design are applicable.

The toroidal type of transformer also finds extensive use in telephone practice. A diagrammatic sketch of a common form is shown in Fig. 214. In this unit the core is made either of soft-iron wire or of thin laminations and, as the name implies, the winding, consisting usually of four sections,

is toroidally arranged, thus giving high efficiency and comparative freedom from stray fields. All eight terminals are brought out to separate connections, thus making a flexible unit. This type of transformer is used in both talking and ringing circuits. Such transformers are frequently spoken of as "repeating coils."

As implied above, the transformers involved in the design of audio amplifiers, used in connection with radio receivers, public-address systems, and phonographic reproduction, constitute a more or less special group. Such transformers must be designed to have a uniform transformation ratio over a wide range of frequencies. If musical sounds are involved, the frequency limits will lie between about 35 and 12,000 cycles/sec. An examination of Eq. (203) shows that emf developed in the secondary of a transformer depends upon the frequency and, indirectly, upon the permeability which, in turn, is a function of the strength of the magnetizing field. Thus it is seen that the design of a multiple-frequency transformer presents a unique problem. This is further complicated by the fact that at the higher frequencies, the electrostatic capacitance that exists between the layers of the winding will by-pass an appreciable part of the current. This capacitance, together with the inductance of the winding, forms a resonant circuit (Sec. 164) for a narrow band of frequencies, thus giving nonuniform voltage transformation

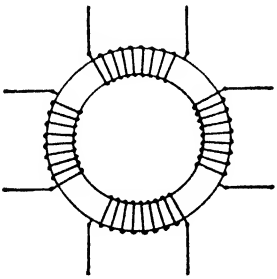


Fig. 214.—Toroidal type of transformer.

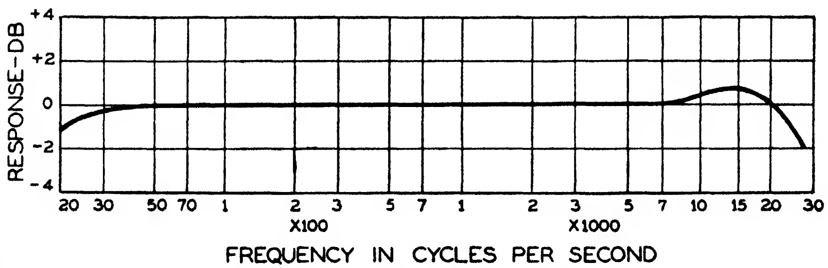


Fig. 215.—Response curve of a high-fidelity audio transformer.

The effect of either process is to give rise to distortion in the wave form. By arranging the windings in alternate sections, the capacitance effect may be minimized. Notwithstanding these several difficulties, design engineers have developed various audio transformers which show a response curve which is essentially flat over the usable audio range. Such a curve is reproduced in Fig. 215. These attainments have been

greatly facilitated by the availability of several of the new magnetic alloys, particularly those having a fairly constant permeability factor. Because of the properties of these special alloys, it is now possible to produce audio transformers of surprisingly small size and weight, and having a frequency response which does not vary more than ± 2 db[1] over a range from 30 to 20,000 cycles. In somewhat larger units the fidelity is ± 1 db; truly a remarkable engineering accomplishment.

In speaking of transformers, it is pertinent to point out a fundamental principle that finds application wherever transformers are utilized in communication circuits.

It may be shown that, in order to secure maximum transfer of energy and to avoid distortion, the impedance of the primary of a transformer used in telephonic or radio circuits **must be equal to** (or match) the impedance (Sec. 161) of the line or other device from which it receives energy. Further, the impedance of the secondary should also match the impedance of the circuit to which it is to transfer energy. Assume, for instance, that a transformer is needed for the purpose of coupling two circuits, one of which has an impedance of 500 ohms, to another whose impedance is 2,000 ohms. A transformer should be employed whose windings have impedances of 500 and 2,000, respectively, the low-impedance side of the transformer being connected with the line of low impedance and the high-impedance side to the high-impedance circuit.

In radio-communication work, wide use is made of air-core transformers, particularly in connection with resonant circuits. These will be discussed in a later chapter.

PROBLEMS

1. In the case of a 10-kw 60-cycle 2,300/115-volt transformer, what is the turn ratio? Neglecting the no-load current, what is the primary and the secondary current?

2. A 10-kw 110/220-volt 60-cycle transformer has a combined iron and copper loss of 100 watts. What will be the primary (220-volt side) and secondary currents? What will be the primary current when the transformer is not connected to a load?

3. In the case of a certain transformer, it is found that the core loss is 100 watts and the full-load copper loss 150 watts. If the transformer is to deliver 5,000 watts to the load circuit, what will be the input in watts? What will be the efficiency?

4. A 60-cycle transformer has 1,000 turns in its primary winding. If the unit is to be operated at flux density of 10,000 gausses when connected to 100-volt

[1] As used here, the decibel (abbreviated db) is a unit which expresses the ratio of input voltage to output voltage in the case of a transformer or audio amplifier (Sec. 237).

mains, what must be the cross section of the primary in order to produce a secondary open-circuit voltage of 5,000 volts?

5. If it is assumed that the regulation of the above transformer must not exceed 5 per cent, how many turns must the secondary winding contain?

6. A 60-cycle 2300/115-volt transformer has 1,200 turns in the primary winding. At what voltage should the unit be operated to produce the same flux density in the core when connected to a 25-cycle supply?

7. A single-phase autotransformer is used for the purpose of dropping the voltage from 115 to 75 volts. The load is 500 watts. Neglecting the losses, what will be the current in the two sections of the transformer winding?

8. A transformer is being designed to operate on a 60-cycle supply, and to develop 1,000 volts at the terminals of its secondary. The secondary winding is to have 2,000 turns; the iron is to be worked at a flux density of 5,000 gausses. What must be the cross section of the core?

CHAPTER XX

ALTERNATING CURRENTS

155. Introductory. It has already been pointed out that in so far as electrical power generation is concerned, by far the greater part of the energy involved is in the form of alternating current. The principal advantages of this form of current in power work may be summarized as follows:

1. An a-c system is much more flexible than a d-c supply. A wide voltage range may be had by means of transformers.
2. Line losses may be minimized by transmitting the energy at high potentials.
3. Alternating-current motors are mechanically and electrically simple and require a minimum of operating attention.
4. Alternating current may be readily converted into direct current when that form of energy is needed.

In referring to the use of alternating current, however, it should be kept in mind that in the field of communication alternating currents play a dominant role. The application of a-c theory to this field will, accordingly, also claim our attention.

A graph representing the relation between emf (or current) and time gives what is referred to as the **wave form.** The brief treatment of the theory of alternating currents which follows is based upon the assumption that we are dealing with sinusoidal waves of current and potential. While the emf available from commercial power circuits may differ slightly from a pure sine wave, calculation based on the sine-wave assumption will closely approximate the actual conditions. If, and when, the wave form is **not** simple, as is usually the case in communication work, it is possible to resolve the wave into components that **are** sinusoidal. This can be accomplished analytically by the aid of Fourier's well-known theorem. It is also possible to determine the harmonic content of a complex wave form experimentally. It therefore follows that any reasoning based on the sine-wave postulate may be extended to waves of any form, provided the components constitute an harmonic series. In any complex wave form the lowest frequency present is designated as the **fundamental or first harmonic.** If a component was present whose frequency was **twice** that of the fundamental, it would be

designated as the **second** harmonic; if three times, the **third,** etc. If upper harmonics (harmonics having a frequency higher than the fundamental) are present in the output of an alternator, they will be the **odd-numbered** ones, *i.e.*, the 3d, 5th, etc. even-numbered harmonics are not present because of the manner in which the emf is generated. In h-f work one may, however, encounter both odd and even harmonics. In general, the harmonics above the fundamental have a much smaller amplitude than that of the fundamental. It should be noted that, in those cases in which the wave form is symmetrical with respect to the

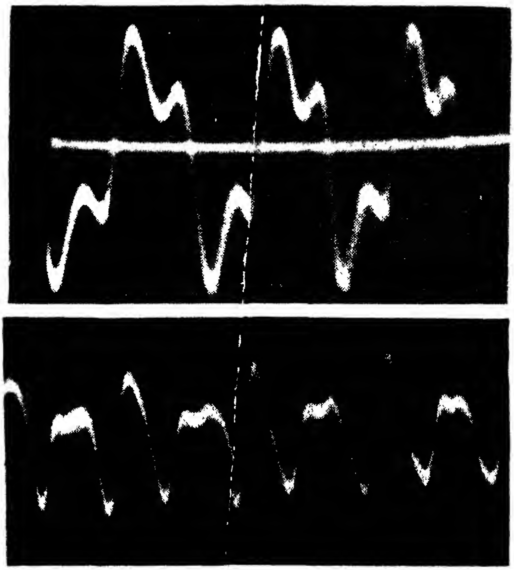


FIG. 216.—Alternating-current (or emf) wave forms. Upper record, symmetrical with respect to zero axis—only odd harmonics present. Lower record, nonsymmetrical—both odd and even harmonics present.

zero axis, **only the odd harmonics are present.** If the wave form shows nonsymmetrical positive and negative loops, both types may be present. These two cases are illustrated in the oscillograms reproduced as Fig. 216.

In power engineering, the frequencies encountered in this country are 60 and 25; abroad 50 and 25 cycles are, or were, common. Owing to the flicker of incandescent lamp, the 25-cycle form of energy is not suitable for illumination purposes. The use of 25 cycles, however, gives a lower line drop than does 60 cycles; and power equipment operates somewhat better at the lower frequency. For these reasons, it has, until recent years, been the custom to use 25-cycle power for electric traction work on railroads and for certain other strictly power purposes. How-

ever, within the past few years certain railroads have changed to 60 cycles and it seems probable that 60 cycles will, in time, become universal in this country.

In communication work, frequencies range from a few cycles per second to many millions. At the higher frequencies, phenomena present themselves which are not commonly encountered at ordinary power frequencies. We shall have occasion to consider certain of these aspects in a later chapter.

156. Effective Value of Alternating Current and Emf. In dealing with alternating currents, the question at once presents itself as to what one means by the terms ampere and volt. In dealing with alternating potentials and currents, we have seen that both the potential and current vary harmonically between zero and a certain maximum value. The relation between the instantaneous and maximum values may be expressed thus,

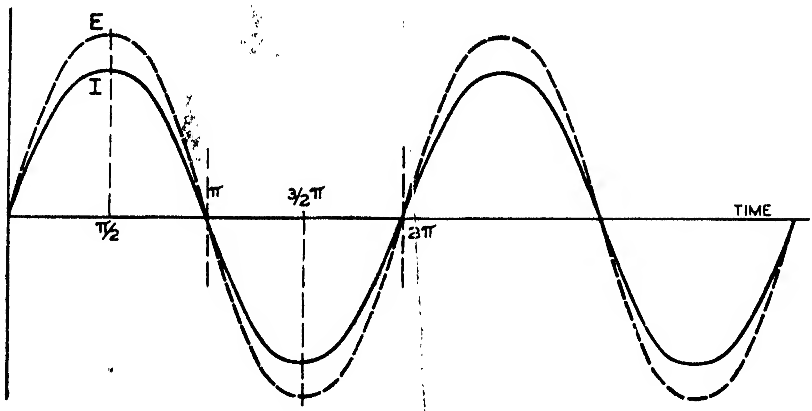


FIG. 217.—Emf and current in phase.

$$e = E_m \sin (\omega t), \tag{207}$$

where e represents the instantaneous value of the emf, E_m the maximum value, $\omega = 2\pi \times$ frequency, and t the time. If such a sinusoidal emf is impressed on a circuit containing only resistance, the instantaneous current will be given by the relation

$$i = \frac{E_m}{R} \sin (\omega t) = I_m \sin (\omega t), \tag{208}$$

where i is the value of the current at any instant and I_m is its maximum value. In such a case, the current maximum will occur at the same time that the emf reaches its greatest value, as indicated in Fig. 217. In the diagrammatic representation as given the relative amplitudes of the

emf and current curves have no significance. Obviously, the maximum value of current and emf cannot serve as the basis of a definition for the ampere and the volt where alternating currents are concerned; and we shall find that the average value is also not a suitable norm. There is, however, a simple and logical basis on which to establish a working definition of an ampere and a volt as it applies to a-c values. Reference is made to the **heating effect of the electric current.** Since the heating effect of the current is independent of the direction in which the electrons are moving, we may take the thermal effect of the current as a basis for our definitions. With this in mind, it may be said that the

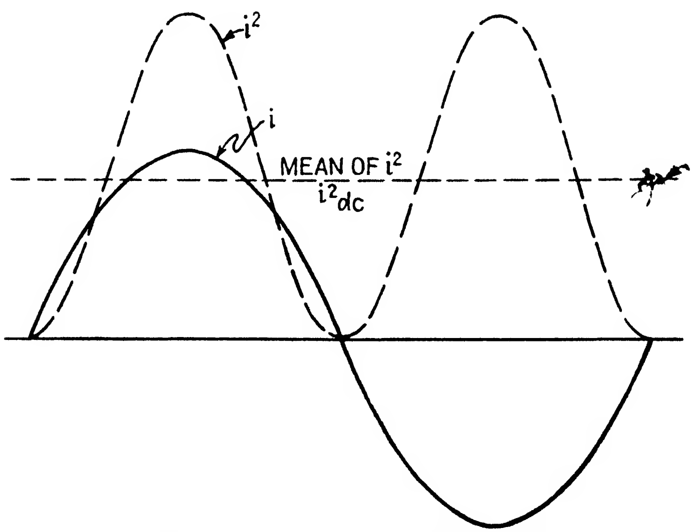


FIG. 218.—Thermal effect serves as basis of effective values of current and emf.

effective value of an alternating current is the value of the direct current that will develop the same thermal effect as the alternating current during one complete cycle of the latter.

It has been shown [Sec. 83, Eq. (110)] that the heating effect of a current passing through a pure resistance varies as the square of the current. Therefore, a curve constructed by using the square of the instantaneous current values as ordinates will represent the variations in the rate at which heat is developed by an alternating current. The mean ordinate of this heat curve also represents **the square of the direct current** that would produce the same quantity of heat in the same period of time as would the original alternating current. This is graphically represented in Fig. 218. It follows, therefore, from our definition of effective value, that the mean ordinate of the curve representing i^2 is

equivalent to the square of the effective value of the alternating current. In other words,

$$I_{\mathrm{eff}}^2 = \text{mean value of } i^2, \tag{i}$$

where I_{eff} is the effective value of the alternating current. Then

$$\begin{aligned} I_{\mathrm{eff}} &= \sqrt{\text{mean value of } i^2} \\ &= \sqrt{\text{mean value of } I_m^2 \sin^2(\omega t)}. \end{aligned} \tag{ii}$$

It now remains to evaluate the expression under the radical sign. By the aid of a simple trigonometric transformation, we may write

$$I_m^2 \sin^2(\omega t) = \frac{I_m^2}{2} - \frac{I_m^2}{2} \cos 2(\omega t).$$

In general, if n is an integer, the mean value of $\cos n\alpha$ over a complete cycle is zero. Therefore, the above expression reduces to $I_m^2/2$, which, from (i), is the square of the effective value of the current. Hence the effective value of an alternating current is given by the expression

$$I_{\mathrm{eff}} = \frac{I_m}{\sqrt{2}} = 0.707 I_m. \tag{209}$$

Since we have assumed that the current and the emf are both sinusoidal, we may write a corresponding expression for the effective value of an alternating emf, thus,

$$E_{\mathrm{eff}} = \frac{E_m}{\sqrt{2}} = 0.707 E_m. \tag{210}$$

From Eqs. (209) and (210) it is evident that

$$I_m = I_{\mathrm{eff}} \sqrt{2} = 1.41 I_{\mathrm{eff}} \tag{211}$$

and

$$E_m = E_{\mathrm{eff}} \sqrt{2} = 1.41 E_{\mathrm{eff}}. \tag{212}$$

Because of the method of determining the effective value of current and emf, the term **square-root-of-mean-square,** or, more simply, **root-mean-square,** is sometimes used to designate what we have called effective value. Some writers use the term **virtual** as the equivalent of effective. In engineering practice root-mean-square (abbreviated rms) is the term most commonly employed. The foregoing discussion leads to the statement that an alternating current that is designated as one ampere is equivalent to a direct current of one ampere, if and **when it develops heat in a given resistance at the same rate as the direct current.** Such a current value is the effective, or rms, value indicated above.

Correspondingly, an alternating voltage which will maintain one rms ampere in a noninductive resistance of one ohm is designated as an effective (or rms) volt. The relative magnitudes of the maximum, effective, and average values of an alternating current are graphically indicated in Fig. 219.

All ordinary a-c instruments are calibrated to read in effective (rms) values. If an a-c voltmeter reads 100, the maximum potential difference under test would be 141 volts.

In the event that the emf or current curve is not sinusoidal, the meter reading will still give rms values. Under these circumstances, the true **maximum** values would not be given by Eqs. (211) and (212). However, those equations would give the maximum values of what is known

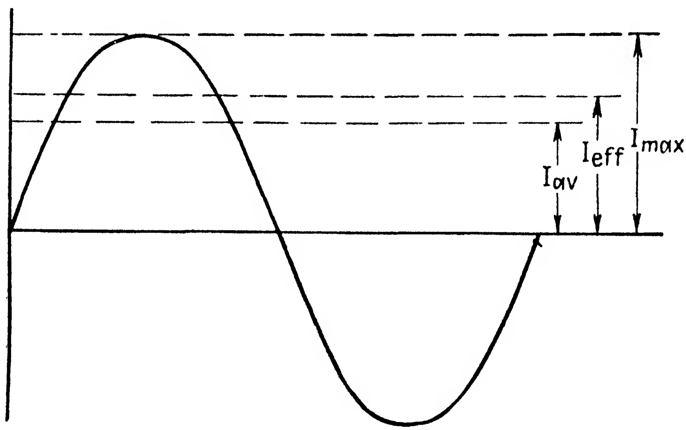


FIG. 219.—Relation of maximum, effective, and average values of current and emf.

as the **equivalent sine curve.** For some purposes it is permissible to use such equivalent-sine-curve values in computations. If more exact values are required, it becomes necessary to break down the complex wave into its components and to deal with each component separately, afterwards combining the several elemental results in order to arrive at the maximum value for the original wave.

From this point on in our discussions I and E without subscripts will always indicate rms values.

157. Average Values and Form Factor. The average value of an alternating emf or current for any number of complete cycles is zero, because there are as many negative values as positive ones. The **mean value for a half cycle,** however, is a quantity that is not zero, and that is useful in certain a-c calculations. To express this in terms of the maximum value of the current, we have but to find the area under the sine

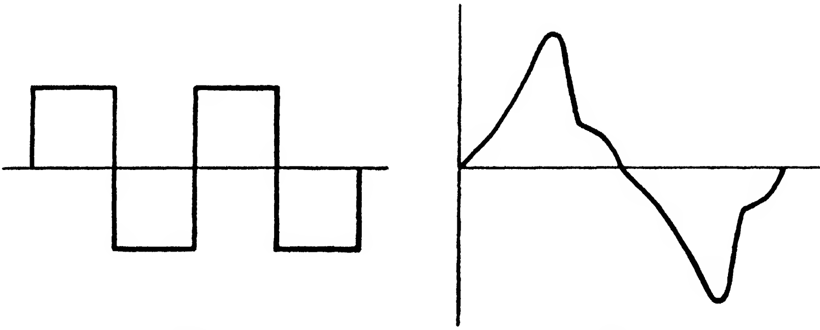
curve and divide it by the length of the base line, expressed as an angle. In terms of the calculus, this would be

$$I_{av} = \frac{I_m}{\pi} \int_0^\pi \sin \omega t \, d(\omega t) = 0.636 I_m. \tag{213}$$

The ratio of the effective to average values is known as the **form factor,** because of the fact that it serves as an indication of the wave form. This ratio is

$$\frac{0.707}{0.636} = 1.11.$$

If the form factor is less than the value above given, the wave form will be flat-topped (Fig. 220*a*); if it is greater than 1.11, the wave is peaked, as shown in Fig. 220*b*. In dealing with the core losses in transformers



(a) (b)

FIG. 220.—Significance of form factor.

and in other a-c equipment, the form factor has a bearing on the magnitude of such losses. If the form factor is high, the core loss will be low, and vice versa. The core loss in a transformer, for instance, when supplied with a current having a flat-topped wave, would show a minimum core loss, while the same unit, when actuated by a current having a form of the type shown in Fig. 220*b*, would have a high core loss. The form factor enters into the consideration of certain problems connected with communication engineering.

158. Vector Representation of Alternating Electrical Quantities. We have seen that both alternating emf and current may be represented by mathematical equations. It has also been shown that these quantities may also be represented by curves plotted on rectangular coordinates. In many cases a third method of representation is found to be useful. Reference is made to what is often called the **rotating-vector method.** Such a

vector is one that is constant in magnitude and rotates about one end at a constant angular velocity. In applying such a plan of representation to alternating electrical quantities as, for example, an emf, the length of the line representing the vector is made to indicate the maximum (or the rms) value of the quantity. The line is thought of as rotating at an angular velocity such that the number of revolutions per second corresponds to the frequency of the emf or current. Reference to the diagram appearing as Fig. 221 will assist in arriving at an understanding of this method of representation.

In the diagram E represents $E_m \sin \omega t$. Whatever the source of the emf may be, its magnitude is assumed to be zero when the vector representing it is in a horizontal position and to be a maximum when

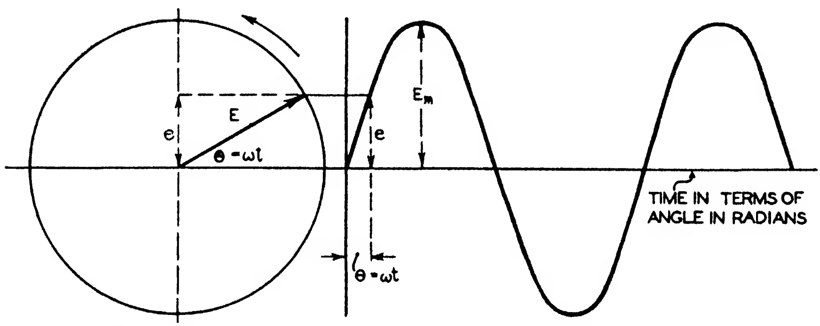


FIG. 221.—Vector method of representing periodic electrical quantities.

vertical. It is further assumed that the vector always rotates counterclockwise. The angle is expressed in electrical angular measure, usually radians. The line representing the vector may be drawn to scale, or, as is more commonly the case, it may be given any convenient length; in which event **phase,** and not magnitude, is the significant aspect of the diagram. The representation shown at the left of Fig. 221 is sometimes referred to as a **clock diagram.**

159. Phase Relations. The space relation of one electrical vector to another is of great importance in a-c computations. As we shall see shortly, if ohmic resistance alone exists in a circuit to which an a-c emf is applied, the current and the emf will be in phase, *i.e.*, both of these quantities (vectors) pass through their maximum and minimum values **simultaneously.** The case is diagrammatically represented in Fig. 222.

Circumstances may be such that the current may not reach its maximum value at the same time as does the emf. In other words, a phase difference may exist between these two vectors. The sketch shown as

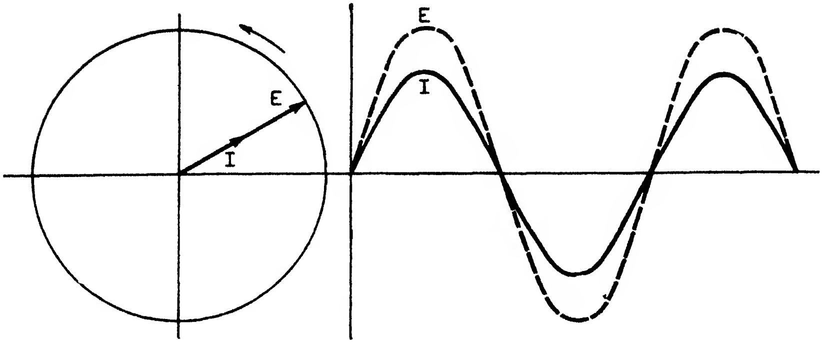


FIG. 222.—Vector representation of the case where the emf and the current are in phase.

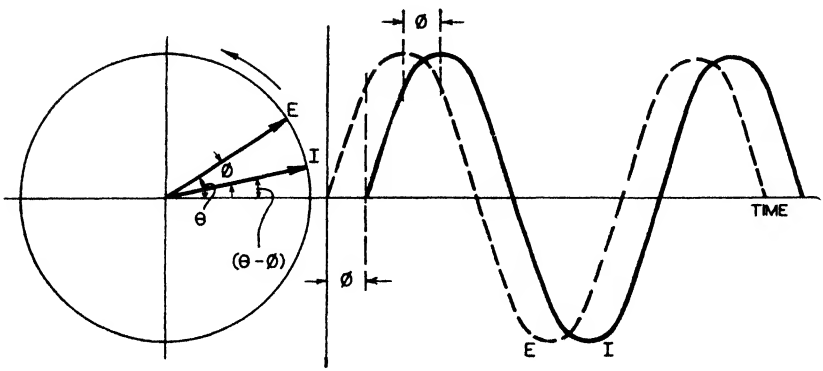


FIG. 223.—Vector representation of the current and emf in an inductive circuit. Note that the current lags the emf.

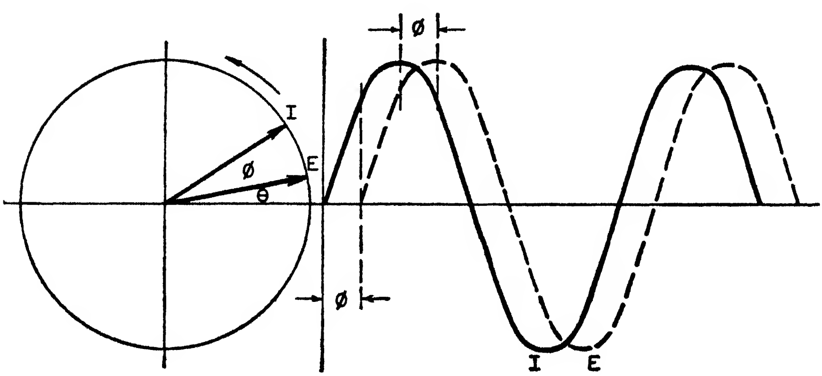


FIG. 224.—Relation between current and emf in a capacitive circuit; the current leads the emf.

Fig. 223 pictures such a situation. The emf would be given by the expression [Eq. (207)]

$$e = E_m \sin \theta,$$

and the equation for the current would be

$$i = I_m \sin (\theta - \phi) \qquad (214)$$

because the current **lags** the emf by the angle ϕ.

Under different circumstances (Sec. 162) the current might reach its maximum value **ahead** of the emf, strange as it may seem. This case is represented by the drawings appearing as Fig. 224. In this case the expression for the current would be

$$i = I_m \sin (\theta + \phi), \qquad (215)$$

the current **leading** the emf by the angle ϕ. It is to be noted that in all three of the cases cited, **the angle used in the equation representing the curve is the angle measured from the zero position of the vector.** The student should clearly understand this important relation.

In practice the curves are usually not drawn; the clock diagrams suffice to represent the phase relations involved. The angle representing the phase difference does not change in value so long as the electrical constants of the circuit remain fixed in value. It is, therefore, immaterial where we place our vectors in the clock diagram.

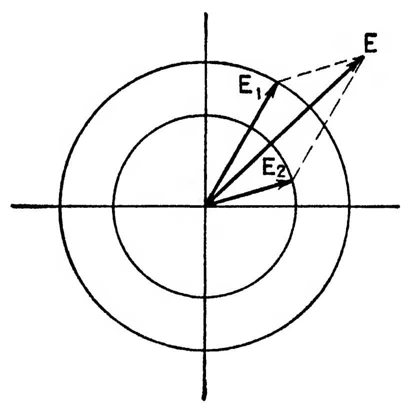


FIG. 225.—Addition of periodic electrical quantities by the vector method.

160. Addition of Alternating Currents and Emf. If and when two emf are simultaneously impressed on a given circuit, it does not follow that the resultant emf will be the arithmetical sum of the two individual emf values. Indeed, in some cases, the resultant emf may be **less in magnitude** than either of the applied emf. The process of vector addition affords a means whereby several emf or current magnitudes may be combined. This is illustrated by the vector diagram shown in Fig. 225. If we assume that the two emf, E_1 and E_2, having the same frequencies, are simultaneously impressed on a given circuit, we may determine, by the usual methods of vector addition, that the vector E will give the magnitude and relative phase relations

of the resultant. The same procedure will also apply when combining two or more current vectors. In any given case it would be necessary to know the phase angles of the components involved. For two vectors the cosine law may be employed in finding the magnitude and phase angle of the resultant; or one may resolve each vector into rectangular components and then combine in the usual manner.

161. Relation between Current and Emf in an Inductive Circuit. We next proceed to derive an expression for the relation between an applied sinusoidal emf and the resulting current in a circuit that contains inductance in addition to resistance, as indicated in Fig. 226. This represents what is known as a **series circuit.** Let the applied voltage be a sine function of the time as given by $e = E_m \sin(\omega t)$, where $\omega = 2\pi f$, f being the frequency. In order to establish a current in such a circuit the applied emf must be equal to the ohmic drop Ri + the counter emf of self-inductance $L(di/dt)$. This may be written in the form

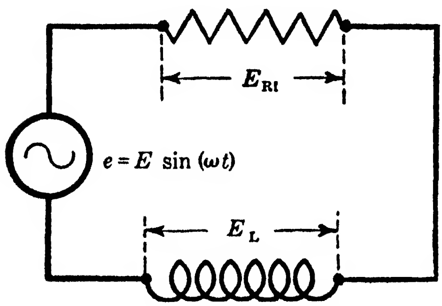


FIG. 226.—A series circuit, containing inductance and resistance.

$$E_m \sin(\omega t) = Ri + L\frac{di}{dt}. \tag{216}$$

This is a linear differential equation of the first order, and the procedure to be followed in arriving at a solution is outlined in any standard text on differential equations. The solution is

$$i = \frac{E_m}{\sqrt{R^2 + \omega^2 L^2}} \sin(\omega t - \phi) + A_1 \epsilon^{-\frac{R}{L}t} + \cdots.$$

After a brief interval of time the exponential terms[1] become negligible. The remaining part of the equation shows that the current is a periodic function of the time and that **it lags the impressed emf by an angle ϕ.**

The greatest value that $\sin(\omega t - \phi)$ can have is unity; hence the maximum value of the current will be given by

$$I_m = \frac{E_m}{\sqrt{R^2 + \omega^2 L^2}}. \tag{217}$$

[1] Immediately after an alternating emf is applied to a circuit there are, in general, in addition to the principal current, one or more other currents of small magnitude that exist for a brief interval of time. These temporary currents are called "transients."

It has previously been shown that the virtual or rms values of both emf and current bear a definite ratio to the maximum values; therefore the above equation holds for virtual or rms values.

The denominator of the right-hand member of Eq. (217) is known as the **impedance** of the circuit, the first term being ordinary resistance and the second the square of two factors, angular velocity and inductance. The term ωL is known as **inductive reactance.** Reactance is commonly indicated by X, in this case X_L.[1]
To rewrite, we have

$$I = \frac{E}{\sqrt{R^2 + X_L{}^2}}, \tag{218}$$

where rms values of current and emf are indicated. Both reactance and impedance are expressed in ohms. It should also be noted that the reactance, and hence the impedance, is a function of the frequency. If the resistance is small compared with the reactance, as is frequently the case in practice, the above relation becomes

$$I = \frac{E}{X_L} = \frac{E}{\omega L} = \frac{E}{2\pi f L}. \tag{219}$$

In Eqs. (217) and (218) the impedance is sometimes represented by Z, thus,

$$Z = \sqrt{R^2 + X_L{}^2}. \tag{220}$$

The foregoing relations will be found to be useful tools. For instance, if one measures the drop over a given inductive winding for known current and frequency values, the magnitude of the inductance may be computed. Knowing the inductance, the current and the frequency, one may compute the "reactance drop," or fall in potential, due to a given inductance. It will be evident that the magnitude of the current in an a-c circuit may, if desired, be controlled by means of a variable inductance rather than by a variable resistance, thus saving the energy that would ordinarily be dissipated as heat if the latter method were employed. An inductance which is utilized for current control in this manner is known as a **choke coil.**

Problem. Suppose we have a circuit whose resistance is 10 ohms and inductance 25 henrys. What will be the magnitude of the current if the applied voltage is 2,000 at 60 cycles? If the frequency was 1,000 cycles?

Solution.

$$\begin{aligned} Z &= \sqrt{10^2 + 4^2\pi^2 \times 60^2 \times 25^2} \\ &= 9{,}425 \text{ ohms.} \\ I &= \frac{2{,}000}{9{,}425} \\ &= 0.212 \text{ amp.} \end{aligned}$$

The student should solve the second part of the problem.

The vector diagram representing the case above discussed is given in Fig. 227. In conformity with Eq. (216), the applied emf OA is resolved into two components, one of which, OB, is to compensate for the ohmic drop RI (which will be in phase with the current), and a second, OC, which must equal the counter emf of self-induction OD. By Eq. (219) the magnitude of the latter component will be given by $I\omega L$. If one takes AB as the geometric equivalent of OC it will be seen that the sides

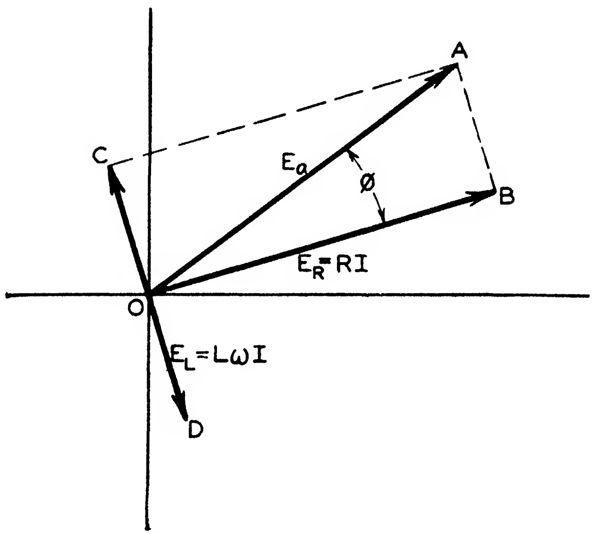


FIG. 227. Vector diagram for the case involving inductance and resistance in series. Note the phase relations.

of the triangle OAB represent the several emf factors involved in the situation under discussion. This vector triangle is redrawn as Fig. 228. By geometry,

$$\overline{OA} = \sqrt{\overline{OB}^2 + \overline{AB}^2}.$$

Substituting, we have

$$E = \sqrt{R^2I^2 + \omega^2L^2I^2},$$

or

$$I = \frac{E}{\sqrt{R^2 + \omega^2L^2}},$$

which is identical with Eq. (217).

The phase difference between the applied emf and the resulting current may be expressed in terms of the circuit constants. From the fore-

going vector relations it will be evident that the magnitude of this phase difference will be given by the relation

$$\tan \phi = \frac{\omega LI}{RI} = \frac{\omega L}{R} = \frac{2\pi fL}{R}. \tag{221}$$

It is to be noted that the angle of lag ϕ varies directly as the self-inductance of the circuit, and also directly as the frequency. It should also be observed that the **resistance** of the circuit enters as a factor in determining the magnitude of the phase angle. Physically the phase difference is **an interval of time** and not an angle; the relation $\phi = \omega t$ serves

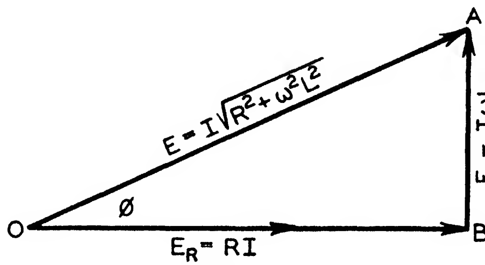


FIG. 228.—Vector triangle for the series circuit.

to connect the time and the angle factors. A practical application of the above phase relation will illustrate its utility.

Problem. Suppose we have a circuit in which there is an inductive winding whose inductance is 0.002 henry and whose resistance is 10 ohms. If the frequency of the current flowing in the circuit is 60 cycles, what will be the angle of lag?

Solution. Making use of the relation embodied in Eq. (221) we have

$$\begin{aligned}\tan \phi &= \frac{2\pi \times 60 \times .002}{10} \\ &= .0758 \\ &= 4°20'.\end{aligned}$$

To express this as a time interval we have

$$\begin{aligned}t &= \frac{\phi \text{ (in radians)}}{\omega} \\ &= \frac{.0756}{2\pi \times 60} \\ &= 0.0002 \text{ sec.}\end{aligned}$$

What would be the angular phase difference if the inductance were doubled? If the frequency were 25?

162. Alternating Current in a Capacitive Circuit. In this case the applied harmonic emf must equal the ohmic drop plus the potential difference developed across the condenser as it acquires a charge. Referring to Fig. 229, and following the same general plan of analytical approach as in the inductive case, we may write for this form of series circuit,

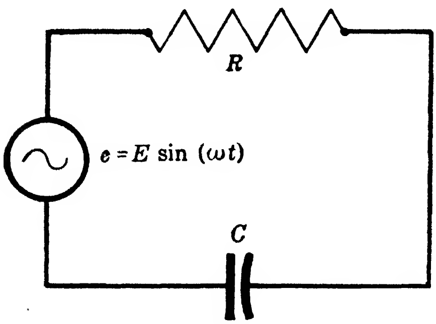


FIG. 229.—A series circuit containing capacitance and resistance.

$$E_m \sin \omega t = Ri + \frac{Q}{C}.$$

But for a charging condenser

$$Q = \int i\, dt;$$

hence we may write

$$E_m \sin \omega t = Ri + \frac{1}{C}\int i\, dt.$$

This equation is also of the first order, and gives as a solution

$$i = \frac{E_m}{\sqrt{R^2 + (1/\omega^2 C^2)}} \sin (\omega t + \phi) + B_1 \epsilon^{(t/RC)} + \cdots .$$

As in the previous case, the exponential terms may be dropped, and we see that the current is again a periodic function of the time. However, it is not in phase with the applied emf, but **leads** [Eq. (215)] by the angle ϕ. The maximum value will be given by

$$I_m = \frac{E_m}{\sqrt{R^2 + (1/\omega^2 C^2)}}. \tag{222}$$

As in the case of the inductive circuit, the denominator of the above relation is known as impedance and the term $1/\omega C$ is designated as **capacitive** (or negative) **reactance,** being written as

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C}. \tag{223}$$

Thus it is seen that the capacitive reactance varies inversely as the frequency and also inversely as the magnitude of the capacitance involved.

Using rms values, Eq. (222) may be written

$$I = \frac{E}{\sqrt{R^2 + X_C^2}}. \tag{224}$$

If the resistance of the circuit is negligible when compared with the reactance, Eq. (224) reduces to

$$I = \frac{E}{X_c} = E\omega C = 2\pi fEC, \tag{225}$$

from which may be derived

$$E_C = \frac{I}{\omega C}. \tag{226}$$

Equation 226 is a very important relation. It shows that the potential difference across the terminals of a condenser carrying a given alternating current varies inversely as the capacitance and inversely as the frequency. Both of these factors are particularly significant in connection with the use of condensers operated at audio and radio frequencies.

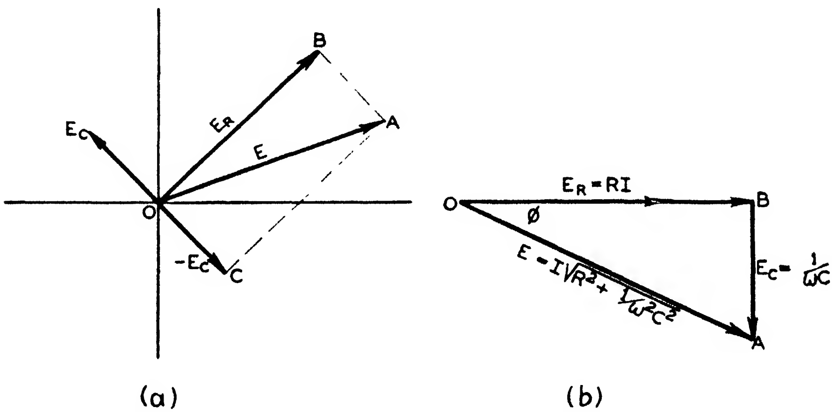


FIG. 230.—Vector diagrams for the case involving capacitance and resistance in series.

The vector diagrams corresponding to this case are to be seen in Fig. 230. They should be carefully studied. The phase angle between the applied harmonic emf and the resulting current may be deduced from the diagram shown as (*b*). It will be given by

$$\tan\theta = \frac{I/\omega C}{RI} = \frac{1}{\omega CR}. \tag{227}$$

By way of illustration, consider the following practical instance.

Problem. Suppose that we have a condenser of 0.01 μf capacitance. Assume $E = 100$ volts and $f = 800$ (mean voice frequency).[1]

[1] The frequency of the alternating current generated by the voice in a telephone circuit has a mean value of about 800 cycles.

Solution. Neglecting the resistance, we have from Eq. (225)

$$I = E\omega C = E2\pi fC.$$

Then

$$I = 100 \times 2\pi \times 800 \times 10^{-6}$$
$$= 0.503 \text{ amp.}$$

If the frequency were 10^6 (common in radio practice) our solution would become

$$I = 100 \times 2\pi \times 10^6 \times 10^{-6}$$
$$= 628 \text{ amp.}$$

It is thus evident that a condenser whose capacitance is of the order indicated above would pass only a small current at voice-current frequency but would readily admit currents of the frequency employed in h-f communication processes.

In charging a long transmission line the current may reach a high value, as shown by the following case.

Problem. Assume a line having a capacitance of 2 μf, which is operated at 100,000 volts and 60 cycles.

Solution. The charging current would be

$$I = 2\pi \times 60 \times 2 \times 10^{-6} \times 10^5$$
$$= 75.7 \text{ amp.}$$

At the voltage mentioned this might, under certain conditions, represent a substantial amount of energy.

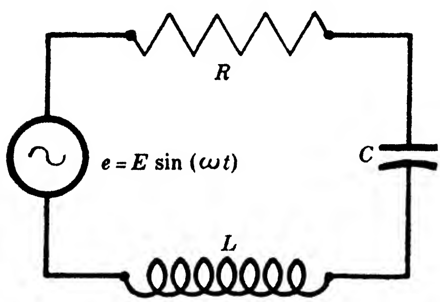


FIG. 231.—A series circuit containing inductance, capacitance, and resistance.

163. Current in a Circuit Having Resistance, Inductance, and Capacitance. A third type of circuit is shown in Fig. 231. Here we have resistance, inductance, and capacitance in series with a source of harmonic emf.

In this case we may express the conditions which obtain by the relation

$$E = E_R + E_L + E_C.$$

This results in the equation

$$E_m \sin \omega t = Ri + L\frac{di}{dt} + \frac{q}{C}.$$

But since

$$i = \frac{dq}{dt},$$
$$\frac{di}{dt} = \frac{d^2q}{dt^2}.$$

Hence

$$E_m \sin \omega t = \frac{q}{C} + R\frac{dq}{dt} + L\frac{d^2q}{dt^2}$$

This is a differential equation of standard form and yields as a solution (neglecting transient terms)

$$i = \frac{E_m}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}} \sin (\omega t - \phi)$$

When $\sin (\omega t - \phi) = 1$ the current will have maximum value and be given by the expression

$$I_m = \frac{E_m}{\sqrt{R^2 + [\omega L - (1/\omega C)]^2}}. \tag{228}$$

This is the most important relation in the theory of alternating currents.

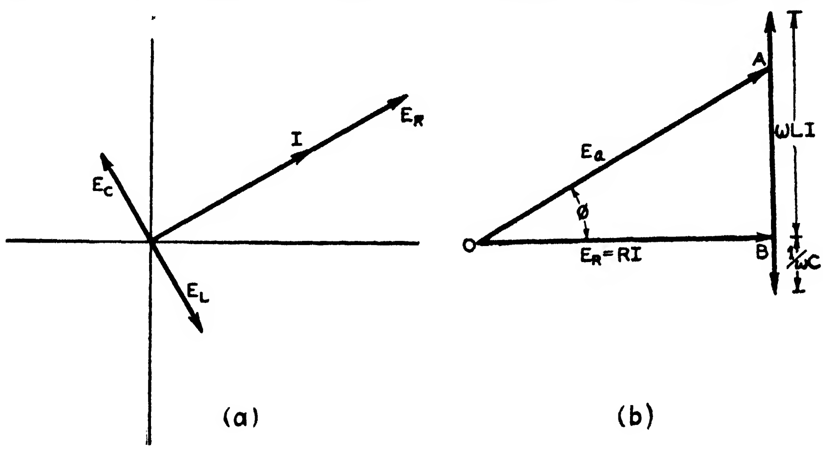


FIG. 232.—Vector diagrams for the case involving inductance, capacitance, and resistance in series.

Using rms values, this equation may be written in either of the two following forms

$$I = \frac{E}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{E}{Z}. \tag{229}$$

As in the two previous cases the denominator of this expression gives the impedance Z, the resultant or net reactance being $\omega L - (1/\omega C)$. Both X_L and X_C are expressed in ohms. The angle ϕ will be either positive or negative, depending upon whether inductance or capacitance predominates. It will thus be seen that a circuit containing both inductance and capacitance may, under certain circumstances, act as if it were

inductive, while under other circumstances it may function as if it were capacitive.

The vector diagrams corresponding to this case are shown in Fig. 232. As seen in (*b*) the inductive reactance exceeds the capacitive reactance by an amount indicated by AB, and hence the current is lagging the applied emf. If

$$X_L = X_C$$

the current would be in phase with the applied emf and, for a given emf, the magnitude of the resistance alone would determine the strength of the current. In the event that the capacitive reactance were to exceed the inductive reactance OA would fall **below** OB, thus indicating that the current was **leading** the applied emf.

Problem. Find the current in a circuit having the following constants: $R = 10$ ohms, $C = 4\ \mu$f, $L = 5$ henrys, $E = 100$ volts, $f = 60$ cycles.

Solution. Substitution in Eq. (228) gives

$$\begin{aligned} I &= \frac{100}{\sqrt{10^2 + \left(2\pi 60 \times 5 - \dfrac{1}{2\pi 60 \times 4 \times 10^{-6}}\right)^2}} \\ &= \frac{100}{\sqrt{100 + (1885 - 648)^2}} \\ &= 0.08 \text{ amp.} \end{aligned}$$

If the condenser in this case were shorted and a constant emf applied to the circuit the current would be 10 amperes.

164. Resonance in a Series Circuit. Referring again to the reactance term $[\omega L - (1/\omega C)]$ of Eq. (228), it is to be noted that low frequencies make the first term small compared with the second, while a high frequency causes the first term to be very large compared with the last. There is, then, an intermediate and definite value of frequency at which the two terms will be equal and this is quite regardless of the particular values of L and C. The particular frequency at which ωL equals $1/\omega C$ may be determined by solving this equality for f thus,

$$\begin{aligned} 2\pi f L &= \frac{1}{2\pi f C} \\ f^2 &= \frac{1}{4\pi^2 LC} \\ f &= \frac{1}{2\pi\sqrt{LC}}. \end{aligned} \tag{230}$$

When the frequency, then, has the value given by this equation, ωL will

equal $1/\omega C$ and we have what is known as a state or condition of **electrical resonance.** In the above equation L is in henrys, C in farads, and f in cycles per second. In practice it is more common to express C in microfarads, and hence our relation becomes

$$f = \frac{1,000}{2\pi\sqrt{LC}} \tag{231}$$

In certain h-f measurements it is convenient to express L in millihenrys. The equation then becomes

$$f = \frac{5,034}{\sqrt{LC}}, \tag{232}$$

C being in microfarads and L in millihenrys.

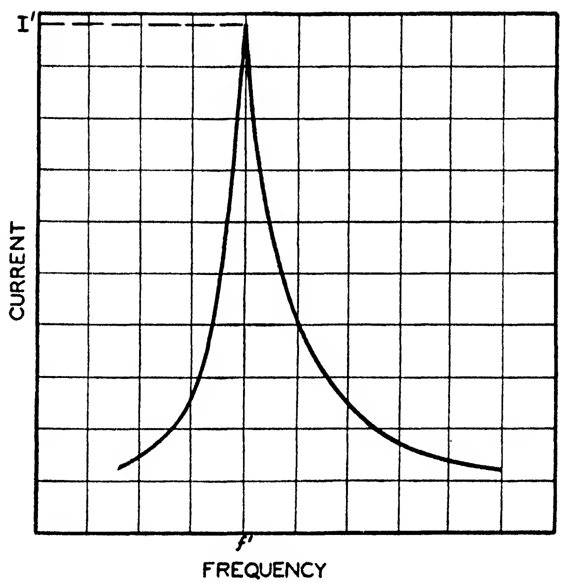


FIG. 233.—Graph showing relation between current and frequency in the case of a series circuit when resonance obtains. f' is the resonant frequency, and I' the corresponding current.

As previously indicated, if a periodic emf having the frequency given in Eq. (230) is applied to a series circuit the magnitude of the current will be determined by R alone. In such a case, if we plot current against frequency there results a graph similar to that shown in Fig. 233, where f' is the resonant frequency and I' the current at resonance. In securing the data for the above curve the capacitance and inductance were held constant, only the frequency of the applied emf being varied. The effect

of resistance is to reduce the value of the current at resonance and also to make the resonance curve less peaked. This is graphically shown by the curves in Fig. 234. At frequencies **below** resonance the circuit reacts capacitively and the current leads; at frequencies **above** resonance the circuit behaves inductively and the current lags.

In the study and application of alternating currents of high frequency the sharpness of the peak of the resonance curve for a given circuit is an important factor. This leads to the use of an expression known as **sharpness of resonance.** It may be shown that sharpness of resonance is given by the ratio $\omega L/R$. It is thus evident that not only the resist-

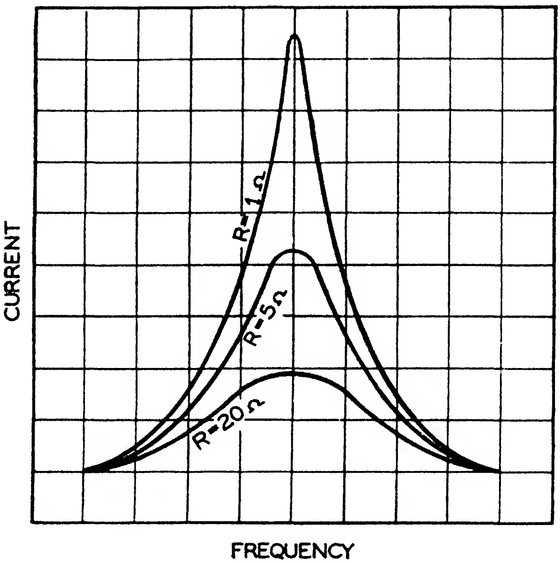


FIG. 234.—Illustrating the effect of resistance on the sharpness of resonance and on maximum current.

ance but also the ratio of the inductive reactance to the resistance are a factor in determining the slope of the resonance curve. Thus, in a circuit containing capacitance, inductance, and resistance, the resonance curve will be comparatively "flat" if the inductance is relatively small and the capacitance high. This is apparent from the curves set forth in Fig. 235. In setting up the two curves shown in the figure the resistance of the circuit was held constant, the frequency alone being varied. The ratio $\omega L/R$ is commonly represented by the symbol Q and is referred to as the **circuit** Q. A circuit having a high Q value will be highly selective.

In dealing with the question of resonance it is important to note that the potential difference developed across the condenser is equal to that

which appears across the inductance coil and also that the magnitude of the potential difference may be several times the value of the impressed emf. This may be well illustrated by a practical case.

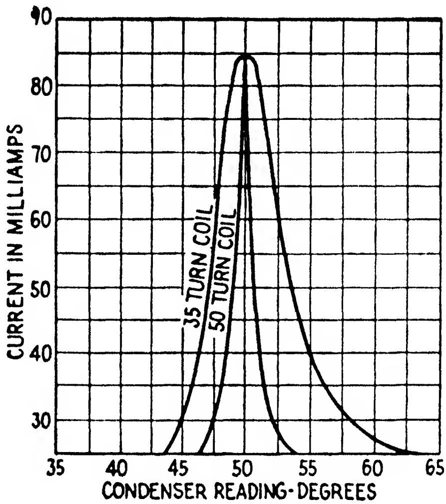


FIG. 235.—Effect of inductance on sharpness of resonance.

Problem. In a certain circuit operated at a frequency of 890,000 cycles, the resistance is 3 ohms, the capacitance 0.0003 μf, and the inductance 0.17 mh. The current at resonance is 6 amp. What is the potential difference across the inductance?

Solution. The value of the impressed emf may be found from the relation $E_a = RI_r$, where I_r is the current when resonance obtains. Substituting known values we have

$$E_a = 3 \times 6 = 18 \text{ volts.}$$

The potential difference across the inductance is given by ωLI_r, or

$$E_L = 2\pi \times 890{,}000 \times 0.00017 \times 6 = 5{,}701.4 \text{ volts.}$$

Since, at resonance, the potential drop across the condenser is equal to that across the inductance, it is evident that both of these units must be insulated to withstand a voltage whose magnitude is more than three hundred times that of the potential impressed on the circuit. Obviously, the condenser, if operated at the frequency indicated, must be constructed so as to withstand this voltage without breakdown.

165. Reactance Curves. In the practical application of the principle of resonance in series circuits various problems arise, in the solution of which certain graphs known as **reactance curves** are found to be useful.

It has already been pointed out that, in the case of a circuit such as

that shown in Fig. 231, the inductive reactance predominates at high frequencies, while at comparatively low frequencies the capacitive reactance is the major factor. In Fig. 236 these two parts of the total reactance are plotted against frequency in terms of ω, the inductive reactance being positive and the capacitive negative. The algebraic sum of these two factors is indicated by the line marked **total** reactance. An examination of the curve shows that at the frequency ω' the inductive reactance equals the capacitive reactance with the result that the **total reactance** is

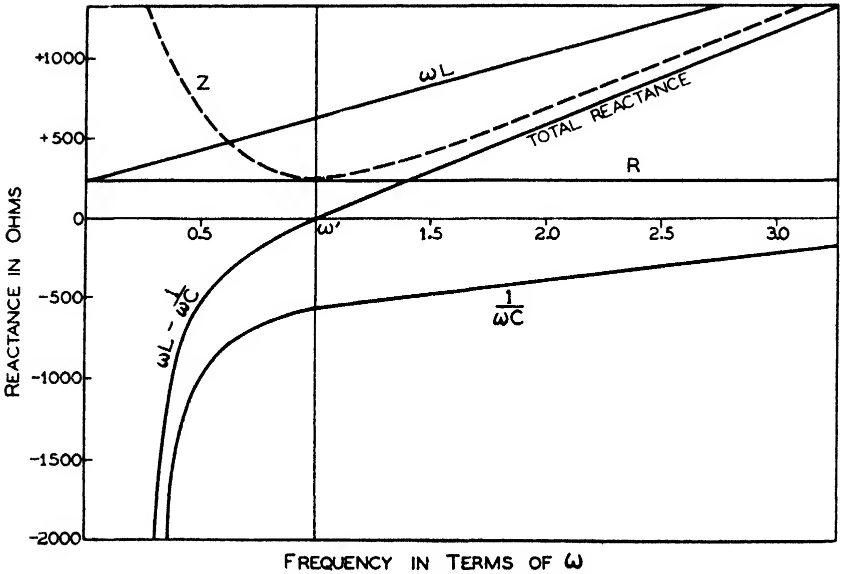


FIG. 236.—Reactance curves for series circuit.

zero. In other words, a condition of resonance obtains and the current in the circuit will be a maximum.

While the total reactance is zero at the resonant frequency, **the impedance is not zero**; therefore the value of the current will be determined by the remaining part of Z, which is R. The values for the impedance Z are shown by the dotted curve.

166. Parallel Alternating-current Circuits. In the a-c circuits thus far studied the source of harmonic emf has been in series with the other circuit elements. We come now to the consideration of a type of circuit in which the applied alternating emf is in **parallel** with the several circuit components. A simple form of such a network is shown in Fig. 237. In the field of applied electricity one encounters parallel circuits more frequently than those of the series type. This is due, in part, to the fact that in the distribution of electrical power a parallel or multiple

circuit system is in wide use. Another reason for the extensive use of this form of circuit is to be found in its frequent incorporation in the filter networks so often utilized in communication engineering.

In the series arrangement the current is the same at all points in the circuit. In the parallel circuit the current through each branch is determined by the reactance of **that particular branch;** and the total current supplied by the source of emf will be **the vector sum of the currents** in the several branches. Mathematically, the above statement may be

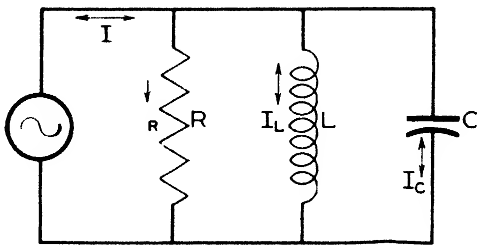


FIG. 237.—Circuit in which the resistance, inductance, and capacitance are in parallel with the source of alternating potential.

expressed thus,

$$\bar{I} = \bar{I}_R + \bar{I}_L + \bar{I}_C, \tag{233}$$

where the bars indicate vector quantities. If we assume that the resistance in the inductive and the capacitive branches is negligible, we may write expressions for the current in the three branches as follows:

$$I_R = \frac{E}{R}$$

$$I_L = \frac{E}{\omega L}$$

$$I_C = \omega CE.$$

The vector diagram representing the conditions that obtain in a parallel circuit would be as indicated in Fig. 238. In dealing with the series circuit we used the current vector as a basis for our diagrammatic representation because it had the same value in all parts of the circuit. For the parallel circuit we can use the applied emf as a basis because, in this case, it is the factor that is the same for all circuit elements. If we assume the resistance of both the inductive and the capacitive branches to be negligible, the current through the inductance will lag the applied voltage by 90°, and the current through the capacitance will lead the emf by a like amount, as indicated in the vector diagram. Hence the current through the inductance will be oppositely directed to that through the capacitance. The current through the resistance will, of course, be

in phase with the applied voltage. If we add I_C and I_L vectorially we get OA'. Combining this vectorially with OD (representing I_R) we have OG, which will then represent the resultant current I supplied by the source of emf. From the geometry of the case, remembering that vectorially I_C is negative with respect to I_L, we may write

$$I = \sqrt{I_R^2 + (I_L - I_C)^2}.$$

By substitution we have

$$I = E\sqrt{\frac{1}{R^2} + \left(\frac{1}{\omega L} - \omega C\right)^2}. \tag{234}$$

It has been shown that the impedance of a circuit is given by the ratio

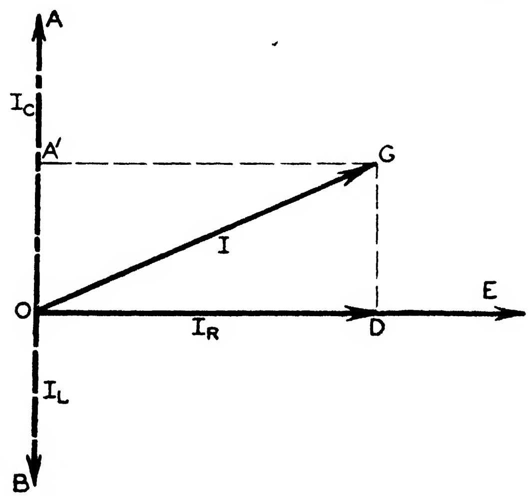


FIG. 238.—Vector diagram representing voltage and current relations in a parallel circuit.

of the impressed emf to the current. Therefore, from Eq. (234),

$$Z = \frac{1}{\sqrt{(1/R^2) + [(1/\omega L) - \omega C]^2}}. \tag{235}$$

In the last equation, the quantity under the radical is known as **admittance** and is designated by the letter Y. It is thus seen that admittance is the reciprocal of impedance, *i.e.*,

$$Y = \frac{1}{Z}, \tag{236}$$

and is therefore the quantity by which the impressed emf must be multiplied to give the magnitude of the resultant current. Admittance $1/Z$ in a-c work corresponds to conductance $1/R$ in d-c practice.

The reactive part of the admittance $[(1/\omega L) - \omega C]$ is called **susceptance.** The first term $1/\omega L$ is commonly designated as **inductive susceptance,** and the last term is known as **capacitive susceptance.** It is to be noted that each of the susceptances is the reciprocal of the corresponding reactance. For an inductive branch the susceptance is considered to be **negative** and for a capacitive branch **positive.** The unit in which conductance, admittance, and susceptance are expressed is termed the **mho,** which is the word "ohm" spelled backwards.

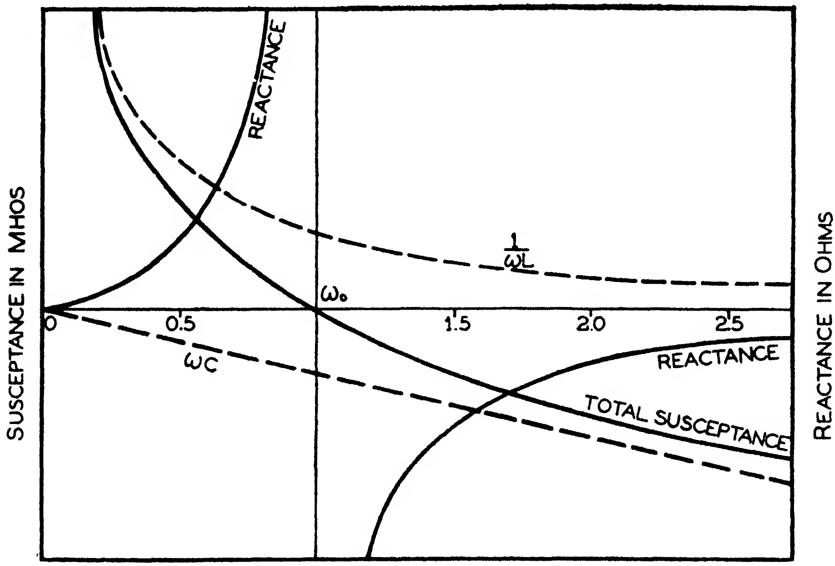


FIG. 239.—Susceptance curves for a parallel circuit.

If we keep in mind that I_L **decreases** with increase in frequency and I_C **increases** with increase in frequency, it will be evident that at some particular frequency the **current** in the inductive branch of a parallel circuit will **equal** the current in the capacitive branch, *i.e.*, $I_L = I_C$. Under those circumstances, the current from the source of emf would be a minimum and equal to I; it would also be in phase with the applied voltage. **When such a condition obtains a parallel circuit is said to be in resonance.**[1]

In a series circuit, at resonance, the inductive and the capacitive **voltages** are equal and opposite, while in the case of a parallel circuit the inductive and capacitive **currents** are equal and opposite. This dif-

[1] This condition is sometimes referred to as **antiresonance** to distinguish it from the corresponding condition in the series circuit when the current is at a maximum.

ference should be carefully noted. At the resonant frequency, if no resistance were present the line current would be **zero.** However, this does not mean that no current is flowing in the inductance and the condenser. Indeed the current in these two branches may be very large—the energy merely passes from the inductance to the capacitance and back again. In short, an oscillating current is set up between these two circuit elements, the applied emf merely supplying the energy to compensate for any circuit losses.

If we plot both the inductive and capacitive susceptance for the parallel circuit being studied we have the curves shown in Fig. 239.

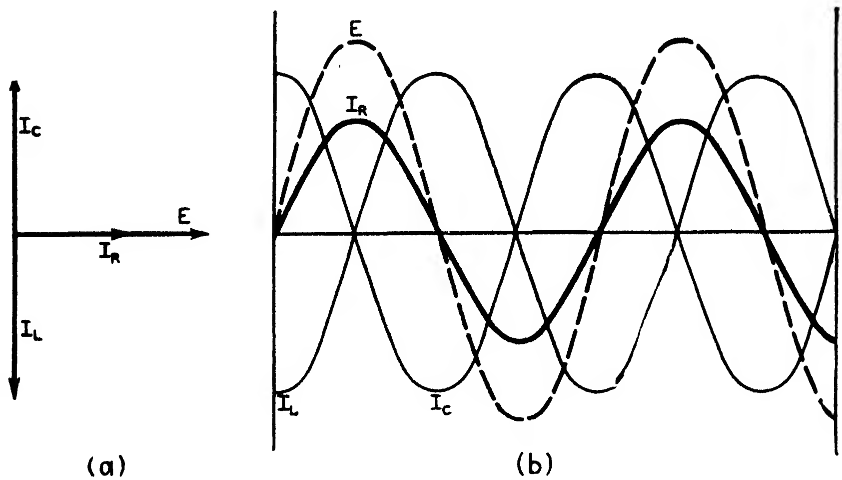


FIG. 240.—Phase relations in the case of a parallel circuit.

By combining the ordinate values for $1/\omega L$ and ωC we may set up a curve that will represent the total susceptance, and it is this quantity that is significant in our present consideration. Bearing in mind the physical significance of susceptance, it is seen that the curve representing total susceptance passes through zero at a frequency ω' which is the resonance frequency for the circuit being studied. This means that the current will be at a minimum when resonance obtains and would be zero if the circuit had zero resistance, thus confirming our previous statement in this respect.

Further, if we set up a reactance curve by taking reciprocals of values on the curve of total susceptance, we find that the curve will have two branches as shown, each of which goes to infinity. It is thus evident that the reactance at resonance frequency is infinite, and that the impedance would also be infinite if the resistance in both branches

of the circuit were zero. Thus we arrive at the same conclusions as previously deduced.

The phase relations of the several currents and the applied emf in the case of a parallel circuit in a condition of resonance are shown graphically in Fig. 240.

Problem. A coil of negligible resistance, whose inductance is 2 mh, is connected in parallel with a condenser across a source of alternating potential whose magnitude is 10 volts at 10^6 cycles. What must be the value of the capacitance in order that a condition of antiresonance may exist?

Solution. Under the condition stated

$$
\begin{aligned}
I_C &= I_L \\
I_L &= \frac{100}{2\pi 10^6 \times 0.002} = 0.008 \text{ amp.} \\
I_C &= 10 \times 2\pi 10^6 \times C = 0.008 \text{ amp.} \\
C &= \frac{0.008}{10 \times 2\pi 10^6} = 1.27\ \mu\mu\text{f.}
\end{aligned}
$$

In many cases each branch of a parallel circuit may involve resistance and some branches may contain resistance, inductance, and capacitance in series. The determination of the current magnitudes, the potential drops, and the phase relations in such cases becomes a more or less complicated problem, and beyond the scope of this text. The interested student will find such cases treated in standard works on electrical and communication engineering.

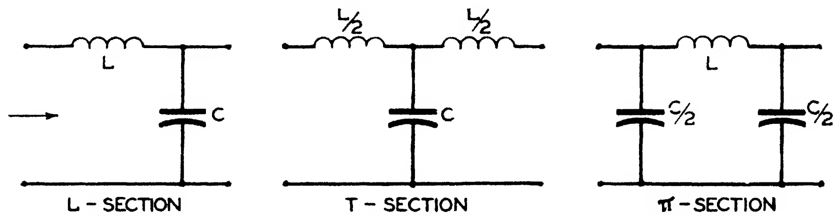
167. Filters. As indicated in the last section, networks consisting of capacitances and inductances arranged in series and parallel combinations are extensively used in the field of communication engineering. In the transmission of speech and music, telegraphic signals, and pictures it becomes necessary to be able to suppress certain frequencies which may be present as components in the original complex electrical wave form. In other cases it is frequently found desirable to eliminate the effects of extraneous electric disturbances. Such results may usually be accomplished by means of some form of electric filter.

Electrical wave filters are four-terminal networks that fall into four general classes as follows

1. **Low-pass filters,** which **suppress** all frequencies above a definite predetermined value, known as the cutoff frequency, and **transmit** the remaining low frequencies.
2. **High-pass filters,** which **transmit** all frequencies above the predetermined cutoff frequency and **suppress** all frequencies below that value.

3. **Band-pass filters,** which are designed to **transmit** a limited band of frequencies and suppress all frequencies that lie outside of that band.
4. **Band-elimination filters,** which **suppress** all frequencies within a limited band and transmit the frequencies that lie outside of that band.

The networks that are utilized for the purpose of accomplishing the above-indicated ends commonly assume one of three forms, as indicated

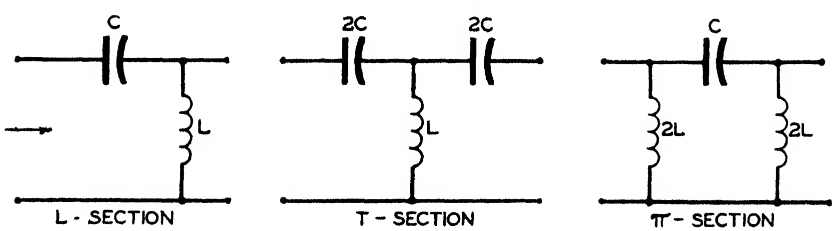


LOW-PASS, SINGLE-SECTION FILTERS

FIG. 241.

in the diagrams appearing in Fig. 241, all of which happen to be of the low-pass type. The names applied to the types of sections are used because of their resemblance to the corresponding letters.

Referring to the simple L type of low-pass filter, it is to be noted that the reactance of the inductance L will be low for low frequencies and high for the higher frequencies. Accordingly, the lower frequencies will suffer little attenuation while the higher frequencies will be greatly



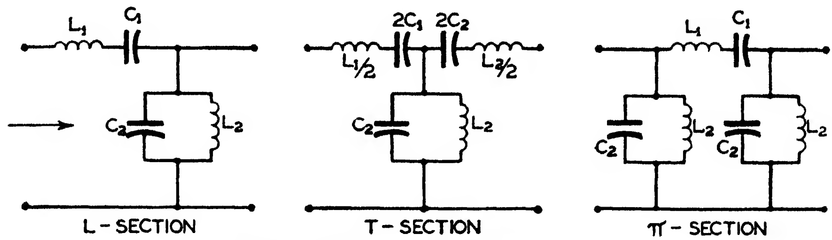
HIGH-PASS, SINGLE-SECTION FILTERS

FIG. 242.

attenuated. This process is aided by the presence of the shunt capacitor. If this capacitance is of low value its reactance will be high at the lower frequencies and low at the higher frequencies; thus the higher frequencies will be by-passed and will not reach the output side of the filter. The components entering into the construction of the other type of low-pass sections function in much the same way as the case just examined.

The high-pass L section (Fig. 242) functions in a manner which is the

reverse of the low-pass case. Here the series low-valued capacitance will offer a low reactance to the currents of high frequency, but will show a high reactance at the lower frequencies. The inductance L will readily pass the low frequencies but will oppose the passage of the higher frequencies, with the net result that the h-f components will pass on to the output side and the l-f currents will be by-passed. The several compo-



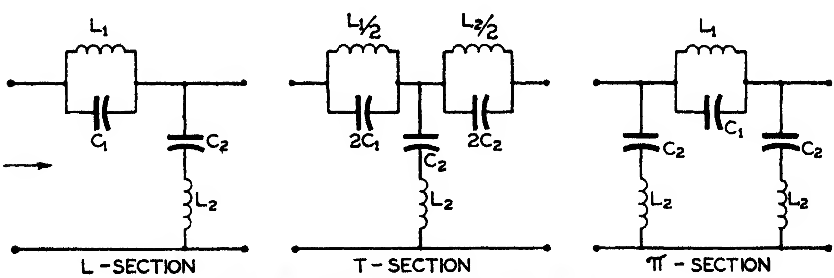
BAND-PASS, SINGLE-SECTION FILTERS

FIG. 243.

nents of the T and π sections of the high-pass units will function in much the same manner.

In Fig. 243 the band-pass arrangement of these simple sections is indicated, and Fig. 244 shows the corresponding band-elimination units.

In designing and using filters, particularly for use at audio frequencies, it becomes necessary to have some method of describing their electrical characteristics. There are several ways of doing this, but the method



BAND-ELIMINATION, SINGLE-SECTION FILTER

FIG. 244.

most commonly used is to express the electrical effects of the filter in terms of **attenuation.** Attenuation refers to the losses which take place in the filter at the various frequencies. The unit of attenuation is the **decibel,** which is indicated by the abbreviation db. The decibel is a unit which is expressed in terms of the logarithm of the ratio of the output to the input, both of which may be expressed in voltage, current, or

power. (This unit will be dealt with later, in connection with our study of amplification.)

The attenuation diagram for the simple low-pass, L-type filter would appear something like (a) in Fig. 245. In (a) and (b) f_c represents the

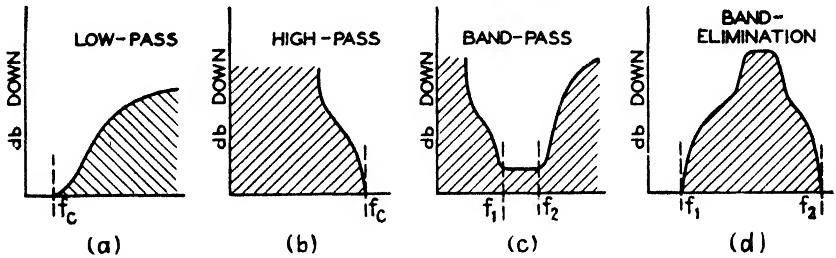


FIG. 245.—Response diagrams for several types of single-section filters.

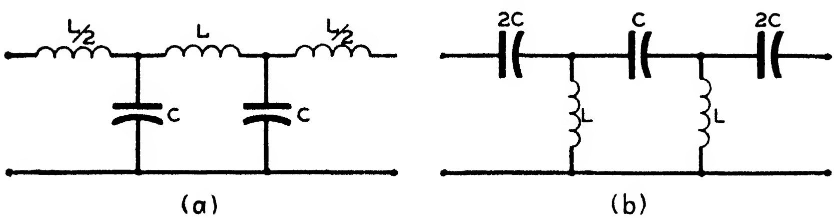


FIG. 246.—A low-pass (a) and a high-pass (b) two-section filter.

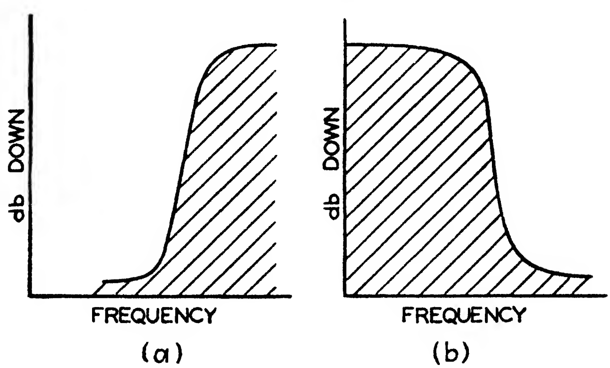


FIG. 247.—Cutoff curves for the two-section filters shown in Fig. 246.

cutoff frequency and in (c) and (d) f_1 and f_2 stand for the frequency limits on either side of the band.

In many cases it is necessary to secure a sharper frequency cutoff than is given by a filter consisting of one section only. Such an end may be attained by assembling a network consisting of several recurrent sections connected in series. For example, a low-pass assembly consisting of two T sections would take the form shown as (a) in Fig. 246. A

high-pass, two-section unit is indicated in (*b*). The corresponding cutoff curves may be seen in Fig. 247. An improvement over the single-section results is to be noted. More than two sections may be used; and, in general, the greater the number of sections the sharper will be the cutoff.

In using a filter one thinks of the input side as connected to a source of harmonic emf, as indicated in Fig. 248. This source (a generator of some type) has a certain impedance. To simplify the case, these impedances have been shown as pure resistances. The output side is connected to a load impedance as shown. Maximum transfer of energy **at the desired frequency or frequencies** will take place when the impedance of the input side of the filter equals the impedance of the generator and the output impedance of the filter equals the impedance of the load. Knowing the cutoff frequency and the impedance in any given case, one may compute the value of the inductance and the capacitance required. For an L-type, low-pass section the two following relations would yield the desired results:

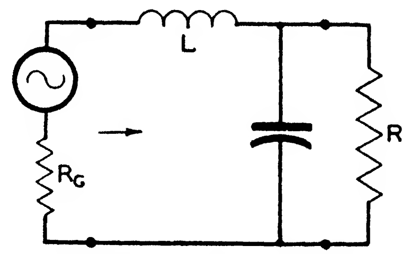


FIG. 248.—Equivalent circuit of a filter when connected between a generator and a load.

$$\left.\begin{aligned} L &= \frac{R}{\pi f_c} \quad \text{henrys} \\ C &= \frac{1}{\pi f_c R} \quad \text{farads.} \end{aligned}\right\} \tag{237}$$

For an L-type, high-pass section

$$\left.\begin{aligned} L &= \frac{R}{4\pi f_c} \\ C &= \frac{1}{4\pi f_c R}. \end{aligned}\right\} \tag{238}$$

For a T-type, low-pass section

$$\left.\begin{aligned} L &= \frac{R}{\pi f_c} \\ C &= \frac{1}{\pi f_c R}. \end{aligned}\right\} \tag{239}$$

For a T-type, high-pass section

$$\left.\begin{aligned} L &= \frac{R}{4\pi f_c} \\ C &= \frac{1}{4\pi f_c R}. \end{aligned}\right\} \tag{240}$$

The foregoing relations will serve to indicate the manner in which one may proceed to compute the constants for a given filter unit. The numerics involved may be illustrated by the solution of a specific example.

Problem. Suppose it is desired to assemble a single-section, T-type filter that will substantially attenuate all currents having frequencies below 200 cycles. Let it also be assumed that the filter is to receive energy from a circuit whose impedance is 500 ohms, and that the unit is to transfer the energy which it passes to a load circuit whose impedance is also 500 ohms. What must be the value of the capacitance and the inductance?

Solution. A high-pass filter will be required. By Eq. (240)

$$C = \frac{1}{4\pi 200 \times 500} = 1.26 \times 10^{-6} \text{ farads}$$
$$= 1.26\ \mu\text{f}$$
$$L = \frac{500}{4\pi 200} = 0.198 \text{ henry.}$$

Filters sometimes assume complicated forms. In dealing with filter calculations, reactance diagrams similar to those shown in Figs. 236 and 239 are found useful. For a detailed treatment of filter theory the reader should consult "Transmission Networks and Wave Filters" by T. E. Shea, or Chap. VI in "Communication Engineering" by W. L. Everitt.

168. Power and the Power Factor. In the case of constant current circuits the value of the power (time-rate of energy dissipation) is given by the product of the current and the emf. In the case of alternating currents, however, that simple relation does not account for all of the factors involved in the situation. In the latter case both the emf and the current vary harmonically, and they may also differ in phase. Thus, if the emf is represented by

$$e = E_m \sin (\omega t),$$

the current would, in general, be given by

$$i = I_m \sin (\omega t - \phi),$$

ϕ being the phase difference between the emf and the current. The instantaneous rate of doing work will be given by

$$P = ei = E_m \sin (\omega t) \times I_m \sin (\omega t - \phi),$$

which reduces to

$$P = \tfrac{1}{2} E_m I_m [\cos \phi - \cos (2\omega t - \phi)].$$

During a complete cycle $\cos (2\omega t - \phi)$ will have all values between $+1$ and -1, and hence its mean value will be zero. The above relation will

accordingly reduce to

$$P = \tfrac{1}{2}E_m I_m \cos \phi.$$

Numerically, however,

$$\frac{E_m I_m}{2} = \frac{E_m}{\sqrt{2}} \times \frac{I_m}{\sqrt{2}}.$$

But we have seen [Eqs. (209) and (210)] that $I_m/\sqrt{2}$ and $E_m/\sqrt{2}$ represent effective or rms values of current and emf, respectively; hence we have

$$P = EI \cos \phi \tag{241}$$

as an expression giving the power P, or time rate of energy dissipation, in terms of the rms values of the alternating voltage, current, and the angle of phase difference. The term $\cos \phi$ is known as the **power factor.** If it were found, for instance, that the current in a given circuit lagged (or led) the emf by an angle of 37°, the power factor would be approximately 0.8, the cosine of 37° being 0.799. In passing it may be said that the angle of lag or lead, and hence the power factor, is largely governed, in any given case, by the nature of the electrical equipment which serves to convert the electrical energy into mechanical energy, *i.e.*, by the character of the electrical load. The value of the power given by Eq. (241) is referred to as the **actual** or **true** power, in contradistinction to the **apparent** power, which is given by the simple product of the voltage and current. The apparent power may be defined as the rate of doing work when the power factor is unity or, in other words, when the current and emf are in phase. The power factor is sometimes defined as the ratio of the actual or true power to the apparent power, thus,

$$\text{Power factor} = \frac{\text{true power}}{\text{apparent power}}.$$

True power is expressed in watts or kilowatts, while apparent power is expressed in terms of the units which go to make up the quantity, viz., "volt-amperes" or, more commonly, "kilovolt-amperes." In practice this is frequently abbreviated kv-a.

The significance of the power factor can be well illustrated by a consideration of the following typical case. Suppose we have a load requiring 10 hp supplied from electrical mains at 220 volts and that conditions are such that the power factor is 0.9. Under these circumstances the current supplied by the service mains would be given by

$$I = \frac{7{,}460}{220 \times 0.9} = 37 \text{ amp.}$$

Suppose that operating conditions so changed that the power factor dropped to 0.6. Computation will show that the current would then be 56.5 amp. It is thus evident that the service mains, and other associated circuits, would carry a much larger current for the same power transfer when the power factor is low. Since the energy lost in the wiring due to heating varies as the square of the current it will be apparent that the losses will be decidedly greater at the lower power factor. To prevent this, and also to obviate other undesirable effects due to a low power factor, public-utility companies take steps to keep the power factor of their service as near unity as possible.

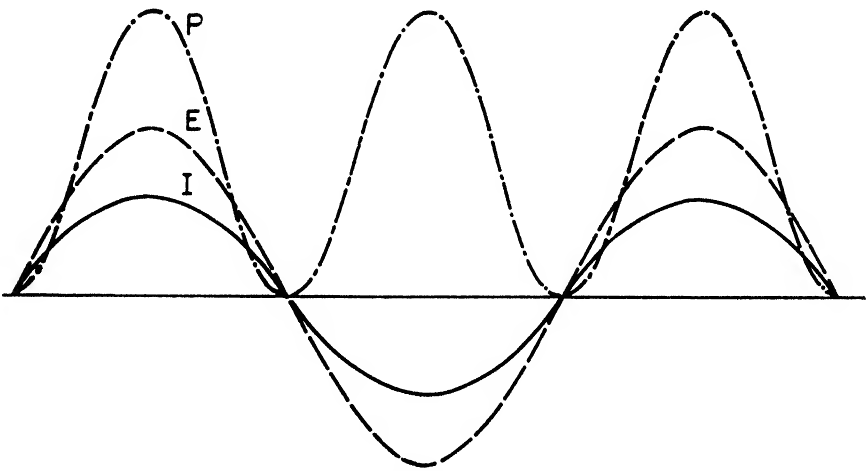


Fig. 249.—Power in an a-c circuit when the current and the emf are in phase.

The important relations discussed above may be given useful graphic form. In the event that the voltage and the current are in phase (zero power factor) Eq. (241) becomes $P = EI$, and we have the situation pictured in Fig. 249. Any given ordinate on the power curve is the product of the corresponding I and E ordinates. It is to be noted that the power varies harmonically and at a frequency twice that of the voltage and current. An alternator whose power factor is zero, operating into a resistive load, would conform to the situation illustrated.

If, however, the electrical load were partly resistive and partly inductive, the current would lag the voltage and the situation could be represented as in Fig. 250. The two shaded areas inclosed by the power curve and the axis represent the energy delivered to, and received from, the load circuit during a complete cycle. The areas marked positive represent the energy delivered **to** the load circuit and those marked negative the energy **returned** by the load to the generator. The **net** energy

expended during a cycle is equal to the algebraic sum of the positive and negative areas.

Theoretically, it is possible to have a condition under which the phase angle would be 90°, as for instance in the case of a circuit containing inductance but no resistance. In such a case the power factor would

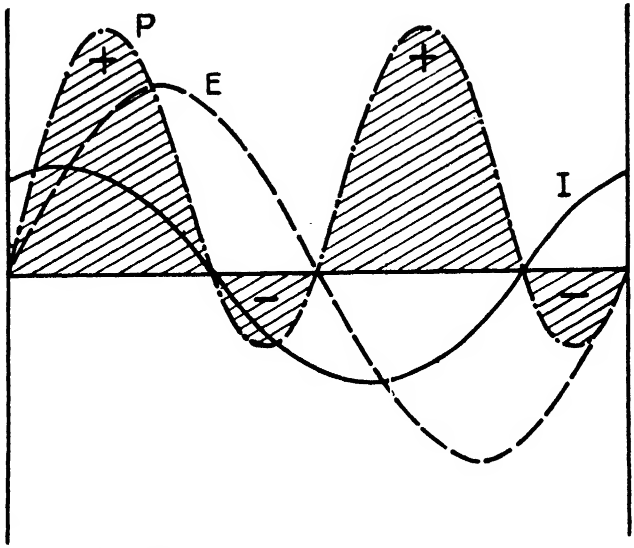


FIG. 250.—Power in an inductive a-c circuit.

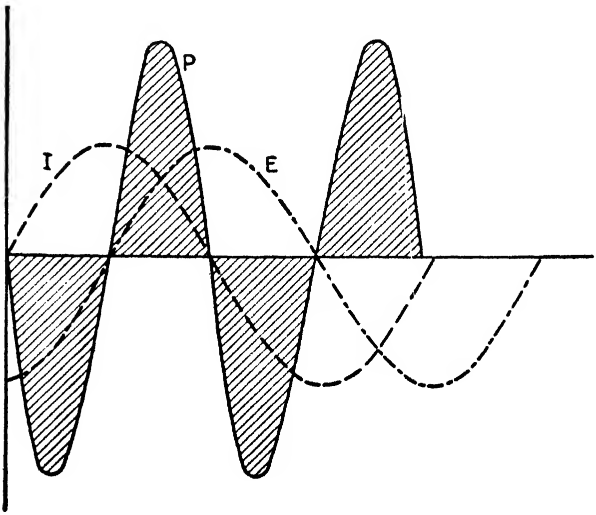


FIG. 251.—Power in an inductive a-c circuit when the current and the emf are in quadrature.

be zero, and the current would be in quadrature with the voltage as shown in Fig. 251.

The diagram shows that the net power is zero, since the sum of the positive areas, for any whole number of cycles, will equal the negative areas. There is current flowing in the circuit under such a condition, but **this current does not accomplish any work.** It is therefore known as the **wattless current.** During one-half a cycle the wattless (or quadrature) current builds up a magnetic field about the inductive winding and during the succeeding half-cycle this magnetic field collapses and the

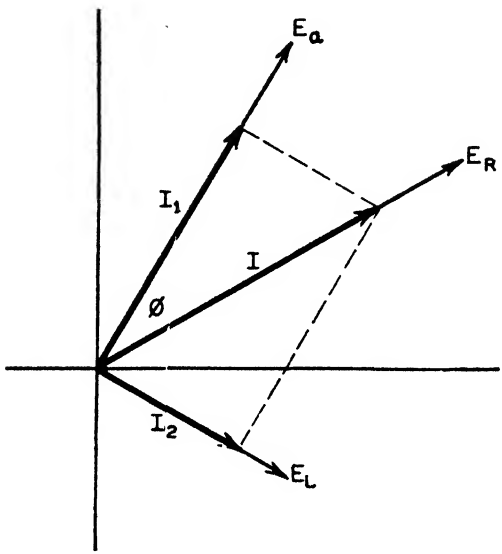


FIG. 252.—Vector diagram corresponding to the case where the current lags the emf.

energy represented by the magnetic field is returned to the original source of emf.

Owing to the fact that all circuits have some resistance, the extreme condition referred to above is never met with in practice. However, there is a **wattless component** in all cases where the phase angle is not zero.

If one considers the case of a circuit in which a lagging current obtains, as indicated in Fig. 252, the current I may be resolved into two components, one of which, I_1, will be in phase with the applied emf, and a second, I_2, will be in quadrature with I_1 and the applied emf. Now $I_1 = I \cos \phi$ and $I_2 = I \sin \phi$. The component $I \cos \phi$ is the useful or **energy component,** while the component $I \sin \phi$ is the so-called **wattless** or **idle component,** sometimes referred to as the **reactive component.**

This latter component represents that part of the current which, in an inductive circuit, gives rise to the magnetic field.

It should not be inferred that two distinct currents exist in the circuit. There is, of course, only one actual current in the line, but the two components of this current give rise to different effects in the circuit. Indeed, the reactive component of the line current produces a loss in the line but does not contribute to the useful work done by the current as a whole.

PROBLEMS

1. The effective value of the difference in potential between two parts of an a-c circuit, as shown by a voltmeter, is 220 volts. What is the maximum or "peak" voltage to which the insulation will be subjected?

2. The ammeter in an a-c circuit reads 10 amp. In any given cycle what are the maximum and minimum values?

3. The self-inductance of a choke coil is 30 henrys. What is its reactance at 60 cycles? At 1,000 cycles?

4. The resistance of a certain series circuit is 12 ohms and the self-inductance is 2 henrys. If a 60-cycle alternating emf of 120 volts is applied to the circuit, what will be the value of the current?

5. What will be the angle of current lag in the case cited in Prob. 4? What will be the power factor?

6. The capacitance of a condenser is 2 μf. What is its reactance at 60 cycles? At 1,000 cycles?

7. The resistance of a series circuit is 5 ohms; its capacitance is 4 μf. If a 25-cycle alternating emf of 110 volts is applied in series, what will be the value of the current?

8. In the case cited in Prob. 7, will the current lead or lag the voltage? What will be the value of the phase angle?

9. An inductance has a value of 20 henrys and carries a current of 0.5 amp. What will be the potential difference developed between its terminals if the frequency of the current is 60 cycles? 1,000 cycles?

10. A circuit that includes a series condenser carries a current of 10 amp at a frequency of 850 kilocycles. The value of the capacitance if 0.01 μf. What potential difference is developed between the terminals of the condenser? What would the potential difference be if the frequency were 500 kilocycles? What would the potential difference be at the two frequencies mentioned if the capacitance was 0.001 μf?

11. In a series circuit the resistance is 5 ohms, the coefficient of self-inductance 2 henrys, and the capacitance 6 μf. What emf, at 60 cycles, must be applied to the circuit in order to maintain an rms current of 10 amp?

12. In Prob. 11, what will be the potential difference at the terminals of the resistance; of the inductance; of the capacitance?

13. What value of capacitance must be connected in series with an inductance of 60 microhenrys in order that the circuit may be resonant at a frequency of 1,250 kilocycles?

14. When a 60-cycle current of 0.1 amp is flowing through a certain inductance, the potential difference at its terminals is found to be 72 volts. If the resistance of the winding is 10 ohms, what is the inductance of the coil?

15. What would be the inductive susceptance of the inductance indicated in Prob. 14?

16. What is the capacitive susceptance at 60 cycles of a condenser having a capacitance of 2 μf?

17. A circuit consists of a resistance of 10 ohms, an inductance of 2 henrys, and a capacitance of 4 μf, in parallel. What is the impedance of the combination at 60 cycles?

18. What is the admittance of the circuit specified in Prob. 17?

19. What will be the magnitude of the total current flowing in the connecting wires when a 60-cycle emf of 100 volts is applied to the network indicated in Prob. 17?

20. It is desired to determine the power factor in an inductive circuit. A wattmeter connected in the circuit reads 500 watts. An ammeter in the circuit shows a reading of 6 amp, and the applied voltage is 115. Compute the power factor.

21. It is desired to suppress all current components having frequencies above 3,500 cycles in a complex wave form by the use of a *T*-type filter. If the filter is to operate between a 500-ohm source and a 500-ohm load, what must be the values of the inductance and capacitance?

CHAPTER XXI

THE SYMBOLIC METHOD AND ITS APPLICATION TO ALTERNATING-CURRENT PROBLEMS

169. Algebraic Method of Representing Vectors. In the course of our survey of the elements of a-c theory it was pointed out that Ohm's law and Kirchhoff's law, in their original forms, were not applicable to a-c networks. The solution of problems in connection with the complicated networks encountered in power distribution and communication work would be greatly simplified if those laws could be utilized in such cases. Fortunately it is possible to express a-c quantities in such a mathematical form that both Ohm's and Kirchhoff's laws may be made to serve as analytical tools in this field.

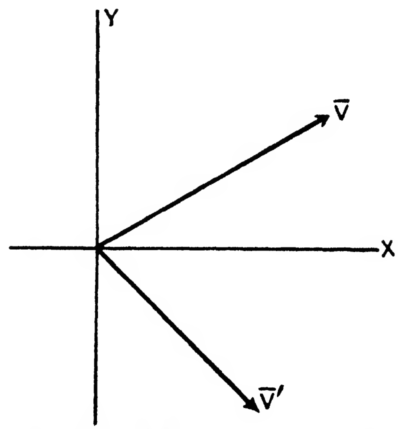


FIG. 253.—Algebraic method of dealing with vectors.

Thus far in our discussion of a-c problems we have made use of the vector diagram method. Is it possible to express vector quantities in such a way that one might deal with these quantities as if they were simple **algebraic** relations? In 1797, Caspar Wessel suggested a mathematical plan by means of which one may deal with both **magnitude** and **direction** in algebraic terms. In 1897, the late Dr. Steinmetz applied Wessel's method in the treatment of a-c problems. In what does this method consist?

Going back to our vector, or clock, diagram, suppose we have any vector quantity, say voltage, as sketched in Fig. 253. This may, of course, be resolved into two components, such as V_x and V_y; and such a vector relation could be written

$$\bar{V} = V_x + V_y, \tag{i}$$

where $\bar{V}$ represents the vector being considered. The right-hand side of the above expression represents a **geometrical** relation, not an algebraic addition. Another way of writing such a relation would be thus,

$$\bar{V} = V_1 + jV_2, \tag{ii}$$

where V_1 represents the component **along the axis of reference** X, and jX_2 the **vertical component.** If such a procedure were followed the factor j would serve to indicate **direction**—a direction at 90°, or in quadrature, to X. If it chanced, for instance, that our vector lay in the fourth quadrant (Fig. 253) the corresponding expression would be

$$\bar{V}' = V_1 - jV_2, \qquad \text{(iii)}$$

the factor $-j$ indicating the perpendicular component **in a negative direction.**

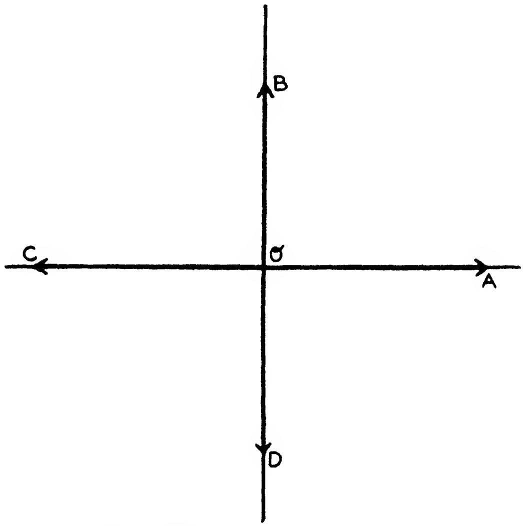


FIG. 254.—Significance of the operator j.

Is it possible to give this direction factor j a numerical significance? Referring to Fig. 254, let $\overline{OA}$, $\overline{OB}$, $\overline{OC}$, and $\overline{OD}$ be vectors of equal magnitude but differing in direction as shown. It will be seen that $\overline{OC}$ differs from $\overline{OA}$ only in sense. In other words $\overline{OA} = -\overline{OC}$, *i.e.*, if $\overline{OA}$ is multiplied by -1 it is rotated through 180°. The factor -1 may therefore be called **an operator** which serves to rotate the original vector through 180°. We have already seen that the factor j represents a rotation through 90°. It therefore follows that $(j)(j)$ or j^2 would represent the 180-degree change in position. But it has just been noted that -1 represents an equal rotation. This leads to the observation that $\mathbf{j^2}$ **must be equivalent to** $\mathbf{-1}$ in its effect on the change of vector position. If, then, $j^2 = -1$, $\mathbf{j} = \mathbf{\sqrt{-1}}$; and this is the value assigned to j by Wessel. Since it is not possible to determine the square root of -1 the term jV in (ii) and (iii) is spoken of as an **imaginary quantity;** V is referred to as a **real quantity.** In dealing with vector quantities the term jV does have a

physical significance; **it is in no sense imaginary.** The expressions indicated in (ii) and (iii) are known as **complex numbers.** That part of any complex number which does not involve j lies in the direction of the X-axis and is called the **reference or in-phase component;** that part of the expression which does contain the factor j is referred to as the **quadrature component.**[1] The magnitude (absolute value) of a complex number is given by the square root of the sum of the squares of the reference component and the quadrature component. In the case of (ii),

$$|V| = \sqrt{V_1^2 + V_2^2},$$

where $|V|$ stands for the numerical magnitude of the vector whose components are V_1 and V_2.

It will be useful to keep in mind the mathematical fact that **any** quantity may be written in the form $a + jb$, where a is a real quantity and jb an imaginary quantity. It is also a fact that if any equation involves both real and imaginary quantities the sum of the real quantities equals zero and the sum of the imaginary terms also equals zero.

From the foregoing discussion it is evident that the quantities encountered in a-c computations may be expressed as complex quantities and their interrelations dealt with by what may be designated as the **symbolic method.** On this basis we may deal with a-c quantities **by algebraic processes;** and thus all the common laws of d-c circuits may be applied to a-c networks.

170. Some Alternating-current Quantities in Terms of the Symbolic Notation. The following list of symbolic equivalents will be found useful in connection with the solution of a-c problems:

Inductive reactance: $+jX_L = +j\omega L$ vector ohms (i)

Capacitive reactance: $-jX_C = -j\dfrac{1}{\omega C}$ vector ohms (ii)

Impedance: $\bar{Z} = R + jX_L$ or $R - jX_C$ vector ohms (iii)

and

$$\bar{E} = (R + jX)\bar{I} \quad \text{vector volts} \tag{iv}$$

Admittance: $\bar{Y} = \dfrac{1}{\bar{Z}} = \dfrac{1}{R + jX}$ vector mhos (v)

and

$$\bar{I} = \frac{E}{(R \pm jX)} \quad \text{vector amperes} \tag{vi}$$

Conductance: $g = \dfrac{R}{R^2 + X^2}$ (vii)

Susceptance: $b = \dfrac{X}{R^2 + X^2}$ (viii)

[1] The X-axis is sometimes referred to as the "axis of reals" and the Y-axis as the "axis of imaginaries."

From (iii) and (iv) above it follows that

$$\bar{Y} = g - jb \qquad \text{vector mhos,} \tag{ix}$$

and

$$\bar{I} = (g - jb)\bar{E} \qquad \text{vector amperes} \tag{x}$$

for the inductive case.

Likewise

$$\bar{Y} = g + jb \qquad \text{vector mhos,} \tag{xi}$$

and

$$\bar{I} = (g + jb)\bar{E} \qquad \text{vector amperes} \tag{xii}$$

for the capacitive case.

From the above notations it may be shown that

$$R = \frac{g}{g^2 + b^2} \qquad \text{ohms,} \tag{xiii}$$

and

$$X = \pm \frac{b}{g^2 + b^2} \qquad \text{ohms.} \tag{xiv}$$

Impedances in parallel: $$\frac{1}{\bar{Z}} = \frac{1}{\bar{Z}_1} + \frac{1}{\bar{Z}_2} + \frac{1}{\bar{Z}_3} + \cdots, \tag{xv}$$

wherein each impedance is expressed in the symbolic (complex) form. Total admittance, when the impedances are in parallel:

$$\bar{Y} = \bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3 + \cdots. \tag{xvi}$$

171. Ohm's and Kirchhoff's Laws in Symbolic Notation. In the complex form Ohm's law may be stated thus,

$$\bar{I} = \frac{\bar{E}}{\bar{Z}}.$$

In similar form Kirchhoff's laws as applied to a-c circuits may be stated as follows

1. **When expressed in symbolic notation,** the sum of the currents at any given point in a circuit is equal to zero.
2. **When expressed in symbolic notation,** the sum of the potential drops in any closed circuit is equal to zero.

172. Application of Symbolic Method to a Typical Alternating-current Problem. Suppose that we have a network as shown in Fig. 255. The magnitudes of the quantities involved are as shown in the diagram. If we assume an applied voltage of 100, what will be the current in each branch and the drop across each branch?

It will be convenient to make the applied voltage the reference vector. On that basis

$$\bar{E} = 100 + j \times 0$$

It will be necessary to find the combined impedance of branches A and B. This will be

$$\begin{aligned} Z_{AB} &= \frac{Z_A Z_B}{Z_A + Z_B} = \frac{(5 - j2)(10 + j5)}{(5 - j2) + (10 + j5)} \\ &= \frac{60 + j5}{15 + j3}. \end{aligned}$$

In order to simplify the above expression we may rationalize the denominator by multiplying both numerator and denominator by unity, *i.e.*, by $(15 - j3)/(15 - j3)$.

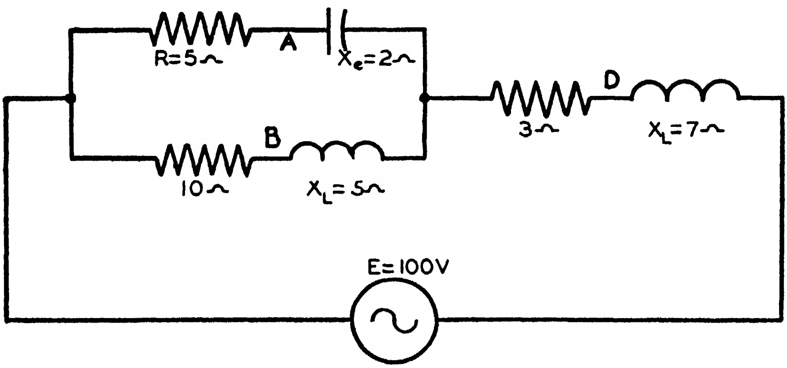


FIG. 255.—Circuit to be dealt with on the basis of the symbolic method.

This gives

$$\begin{aligned} \bar{Z}_{AB} &= \frac{(60 + j5)(15 - j3)}{5^2 - 3^2} \\ &= \frac{61 - j105}{216} = 0.28 - j0.49. \end{aligned}$$

The total impedance of the circuit will be

$$\begin{aligned} \bar{Z}_t = \bar{Z}_{AB} + \bar{Z}_D &= (0.28 - j0.49) + (3 + j7) \\ &= 3.28 + j6.51. \end{aligned}$$

For simplicity we shall write this as

$$\bar{Z}_t = 3.3 + j6.5 \text{ vector ohms.}$$

The total current would be

$$\bar{I}_t = \frac{\bar{E}}{\bar{Z}} = \frac{100 + j \times 0}{3.3 + j6.5} = \frac{100(3.3 - j6.5)}{3.3^2 + 6.5^2} = 22 - 43j \text{ vector amperes.}$$

The actual value of the current would be

$$|I_D| = \sqrt{22^2 + 43^2} = 48.4 \text{ amperes.}$$

The voltage drop across branch D is

$$\bar{E}_D = \bar{I}_D\bar{Z}_D = (22 - 43j)(3 + j7) = 367 + 25j \text{ vector volts.}$$

The absolute value of the drop over branch D would be

$$|E_D| = \sqrt{367^2 - 25^2} = 371 \text{ volts.}$$

The drop over branches A and B is given by

$$\bar{E}_{AB} = (100 + j \times 0) - (367 + 25j) = (-267 - 25j) \text{ vector volts,}$$

and

$$|E_{AB}| = \sqrt{-267^2 - 25^2} = 273 \text{ volts.}$$

Why can the drop across both D and AB be greater than the applied voltage?

It is left for the student to determine the actual values of I_A and I_B and to draw a vector diagram showing the relative position of the several factors involved in the problem.

For a more extended discussion of the complex numbers, as applied to a-c problems, the reader may consult "Electrical Engineering" by C. L. Dawes, Chap. III, Vol. II; or "Alternating Current Circuits" by M. P. Winebach, pp. 53*ff*.

PROBLEMS

All problems in the following list are to be solved by the use of complex quantities.

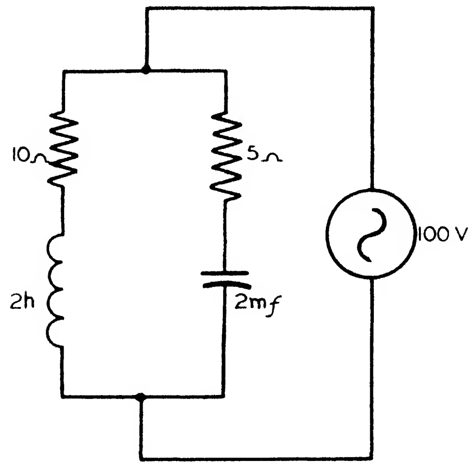
1. A series circuit consists of a resistance (10 ohms) and an inductance (10 henrys). Determine the impedance, the admittance, and the susceptance of the circuit. The frequency may be taken as 60 cycles.

2. In Prob. 1 above, what would be the magnitude of the current if an emf of 100 volts was applied?

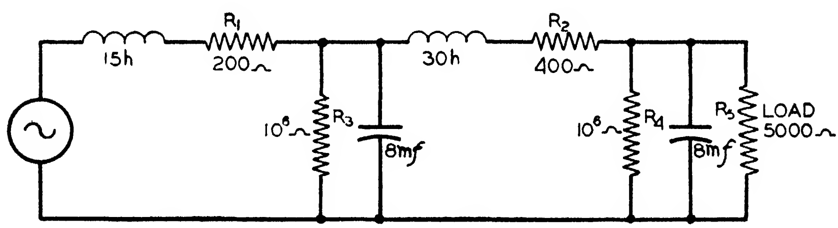
3. A series circuit consists of a resistance (10 ohms) and a capacitance (5 μf). If the operating frequency is assumed to be 60 cycles, what will be the value of the impedance, the admittance, and the susceptance?

4. In the case of Prob. 3 above, what would be the value of the current if an emf of 120 volts were applied?

5. In a circuit such as that shown in the accompanying sketch, what will be the total current and the current in each branch, assuming the frequency to be 60 cycles?



6. As we shall see later, a frequently used type of filter takes the form shown in the accompanying diagram. The inductances are choke coils, and the capacitances are electrolytic condensers. The values given are representative. R_3 and R_4 represent the "leakage resistance" of the condensers. Assuming a frequency



of 120 (the ripple frequency from a rectifier), find the total impedance of the network.

7. Assuming that a voltage of 300 is applied at the input end, in the above case, what will be the drop across the load resistance of 5,000 ohms?

CHAPTER XXII

ALTERNATING-CURRENT INSTRUMENTS

173. Electrodynamometer Type of Instrument. In discussing the D'Arsonval type of instrument (Sec. 119), it was pointed out that such an electromagnetic system would not function as the basis of an a-c meter because with each reversal of the current the torque acting on the moving system would also be reversed, with the result that a vibrating motion would be imparted to the moving coil. Since alternating currents constitute so large a part of the electrical energy utilized, it becomes necessary to arrange for the measurement of alternating current, potential, and power. One widely used type of a-c instrument makes use of the uniform magnetic field set up by what amounts to a pair of Helmholtz coils in place of the field due to a permanent magnet. No magnetic material is thus involved. The magnetic reactions between a pivoted, movable coil and the above-mentioned field coils develops a torque that causes a rotation of the former. Any rotation caused by this torque is limited by coiled springs, as in the D'Arsonval instrument. The magnetic relations are indicated by the sketch shown as Fig. 256. In the diagram S and S' are two stationary coils connected in series. When carrying a current these coils will set up a field whose direction will be as indicated by NS. If current flows through the movable coil M, when in the position shown, the resulting field will have the direction indicated by $N'S'$. Elementary considerations will indicate that the coil M will tend to rotate clockwise and that it will continue to do so until the counter-mechanical torque due to the control springs equals the magnetic torque due to the reaction between the fixed and movable coils. If the two field coils and the movable coil are connected in series, the same current will traverse both sets of windings. Therefore, if the current changes direction in the coils S and S' it will also simultaneously change direction in

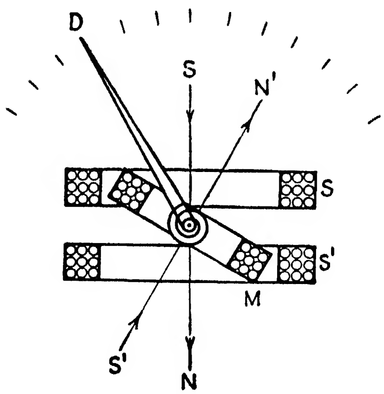


FIG. 256.—Essentials of the electrodynamometer type of instrument.

M. The direction of the magnetic torque will therefore remain fixed. [Actually the magnetic torque is pulsating (120 per second), but due to the damping a definite deflection will occur.] Thus we have available an electromagnetic system the rotation of which is independent of the direction of the current and can therefore be utilized as the basis of an a-c indicating instrument. Since the same current flows through both the fixed coils and the moving coil, and since the magnetic field in both cases is proportional to the current, the **turning moment will be proportional to the current squared.** This means that the lower readings on such an instrument, when used as an ammeter or voltmeter, will be crowded. This is a disadvantage and means that, in practice, a given

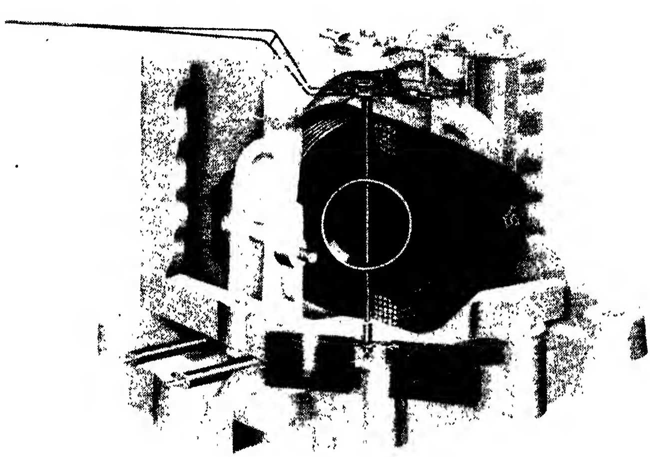


FIG. 257.—Cutaway view showing the several components of an electrodynamometer type of indicating instrument. (*Courtesy of Weston Electrical Instrument Corp.*)

instrument can be used for a limited range only. Since the moving coil operates in a relatively weak field this type of instrument is subject to the effect of stray magnetic fields, unless special magnetic shielding is provided. However, the absence of any permanent magnet makes possible a calibration that is practically permanent. Electrodynamometer instruments are free from frequency and wave-form errors. Since the deflection is proportional to the square of the current, it can be calibrated to read in rms values on alternating current (Sec. 156) and will also give true d-c values. Such an instrument is equally accurate on alternating and direct current.

One form of this type of instrument is made "dead beat" by means of an aluminum damping vane attached to the moving system and moving in a nearly airtight enclosure. One manufacturer uses a horizontal

damping vane which moves over the poles of a permanent magnet, thus effecting magnetic damping. Current is led into and out of the moving coil through two spiral springs, which also function as control elements. The illustration appearing as Fig. 257 shows the essential components of an instrument of the electrodynamometer type.

When the electrodynamometer type of instrument is used as a **voltmeter,** both the field coils and the moving coil are wound with fine wire and are connected in series with a high noninductive resistance. Under these circumstances, when connected across the circuit to be measured, the current in the coils will be proportional to the applied potential difference and hence the scale can be calibrated to indicate either rms or d-c values.

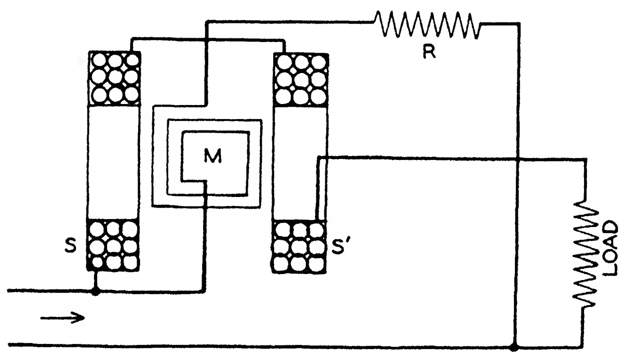


FIG. 258.—Essentials of the electrodynamometer type of wattmeter.

When designed to function as an ammeter, an instrument operating on the electrodynamometer principle has its field coils wound with wire of a size to permit passage of the desired current without undue heating. The current that passes through the field coils also passes through a suitable shunt whose electrical characteristics are similar to those of the moving coil. The winding constituting the moving coil is connected **in parallel** with this shunt. Under these circumstances the current in the moving coil is proportional to the main current; hence the torque is proportional to the square of the current and the instrument can accordingly be calibrated to read direct or rms values.

It is, perhaps, in a **wattmeter** assembly that the electrodynamometer principle finds its widest application. Figure 258 is a schematic diagram of a wattmeter circuit. It is to be noted that the field coils S and S' are connected in series with the load while the moving coil M is connected across the line, through a high, noninductive resistance R. Under these conditions the current through M will be proportional to the applied potential difference. Since the magnetic field due to the current in the

fixed coils will be proportional to the instantaneous current in the line, **the resultant torque will be proportional to the product of the current and the voltage,** and hence to the actual power; and this regardless of the frequency, wave form, or power factor. Wattmeters are calibrated in watts or kilowatts, and read with equal accuracy on either alternating or direct current, if the assembly is properly shielded. A dynamometer type of wattmeter is commonly calibrated on direct current by means of a potentiometer assembly.

174. Iron-vane Instruments. A type of a-c instrument which is widely used in both portable and switchboard models makes use of an

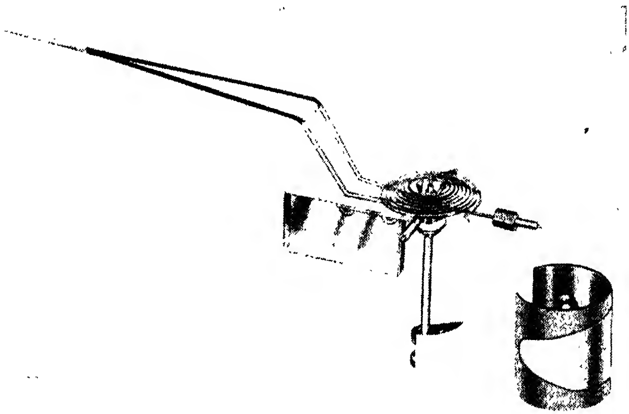


FIG. 259.—Moving system of iron-vane type of instrument. Note the damping vane. (*Courtesy of Weston Electrical Instrument Corp.*)

iron vane as the movable element. A curved rectangular soft-iron vane is attached to a light pivoted shaft, as shown in Fig. 259. A second curved soft-iron member, more or less wedge-shaped in outline, is rigidly supported in the position indicated in Fig. 260*a*. Both the movable and the fixed vanes are positioned within a fixed field coil, as shown. As current flows through the field coils, the upper edges of both the fixed and the movable vanes will acquire the same magnetic polarity. The same will be true of the lower edges. The resulting magnetic repulsion gives rise to a torque that rotates the moving system against the spiral control spring seen at the top in Fig. 259. The value of the magnetic torque, and the resulting rotation, is a function of the magnitude of the current in the field coil, the shape and relative position of the vanes, and the magnetic characteristics of the material constituting the vanes. Due to the fact that the magnetic fields of both vanes are proportional to the current, the deflection of the moving system follows a square law; hence the scale is not uniform, being crowded at its lower end.

When designed for use as an ammeter, the field coil of this type of instrument is wound with heavy wire. When constructed to function as a voltmeter, the coil is wound with many turns of fine wire connected in series with a high resistance, as indicated in Fig. 260*b*. Air damping is provided in both the ammeter and voltmeter units. The instruments are calibrated in rms values, the calibration, however, cannot be carried out with direct current. On alternating currents the iron-vane type of instrument will give an accuracy of at least 0.5 per cent, and can be used on direct currents with a precision of from 1 to 2 per cent. This

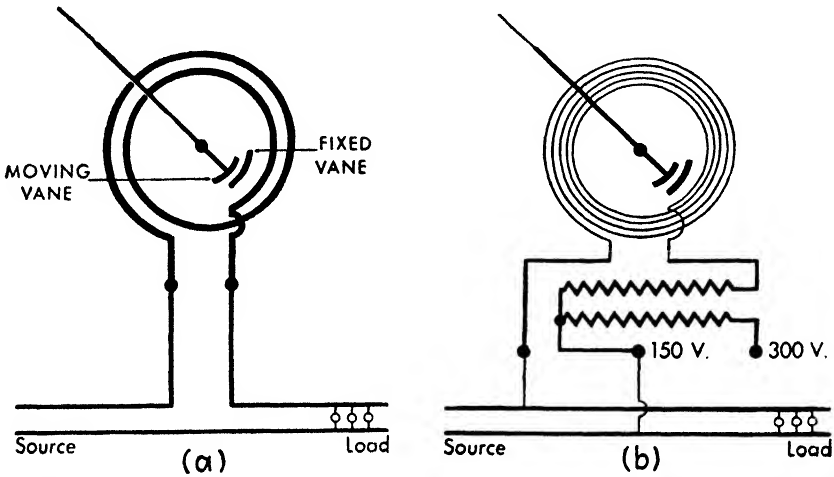


FIG. 260.—Relation of the components in an iron-vane type of meter. (*Courtesy of Weston Electrical Instrument Corp.*)

type of instrument is relatively simple in construction, robust, and comparatively inexpensive to produce.

175. Rectifier Type Meters. Owing to the relatively large amount of power consumed, and the consequent lack of sensitiveness, both the electrodynamometer and iron-vane type of meter are not adapted to the measurement of alternating currents of a few milliamperes, or a-c potential differences of a few volts.

In recent years a type of a-c meter has come into extensive use which utilizes a rectifying unit in association with a D'Arsonval type of indicating instrument, and which does not have the limitations above indicated. The relations of the electrical components are set forth in Fig. 261*a*. Four copper oxide sections are arranged in a bridge circuit, thus providing for full-wave rectification. When the terminal *a* is negative, the electronic current flows through the network by the path *abdc*; when the point *c* is negative the current travels along the path *cbda*; thus the

current through the meter is **unidirectional.** The deflection of the indicating instrument will be proportional to the average value of the alternating current. The instrument, however, is calibrated to read in rms values, and while not strictly linear, the scale is only slightly compressed at the lower end. The rectifier unit is shown, about full scale, in (*b*). The D'Arsonval indicating meter commonly utilized has a full-scale d-c reading of the order of 1 ma. Occasionally the indicating instrument may even be a microammeter. When used as a voltmeter a series resistor is incorporated in the assembly, as shown in (*b*). When the instrument

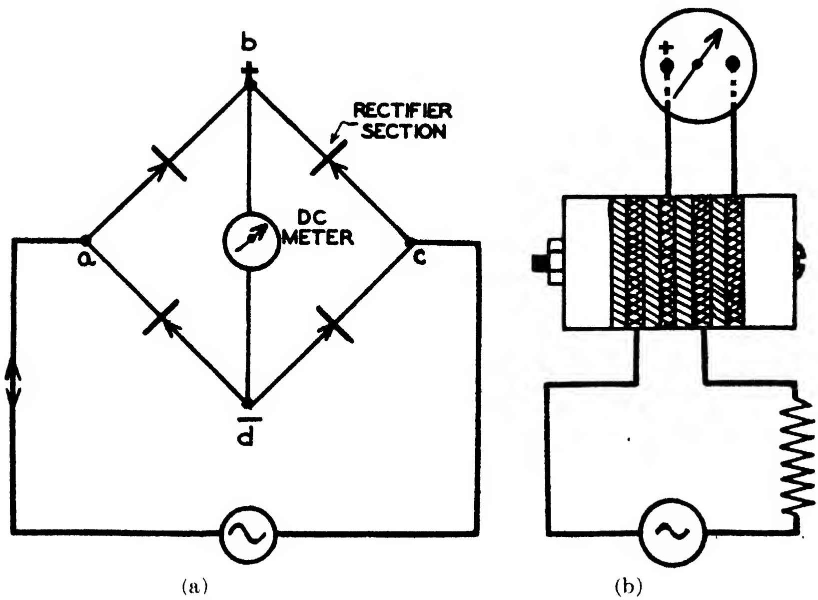


FIG. 261.—Essentials of the rectifier type of meter.

is to function as a milliammeter, a shunt is arranged in the usual manner. When properly calibrated, and not seriously overloaded, this type of instrument has an accuracy of about 1 per cent. The accuracy, however, is effected by temperature and, to some extent, by wave form. Rectifier voltmeters are sometimes calibrated as **volume-level indicators,** or **decibel meters,** for use in connection with audio amplifiers. For a comprehensive discussion of this type of meter the reader may consult an original paper by J. Sahagen in *Proceedings IRE*, February, 1931.

176. Watt-hour Meter. A watt-hour meter is an integrating wattmeter; it measures **electrical energy,** and its readings serve as the basis of electrical merchandizing. There are two general forms of such instruments—the electrodynamometer type and the induction type. By changing slightly the construction of the electrodynamometer form of

indicating wattmeter this type of instrument can be made to function as an energy meter. The modifications involve the replacement of the spiral springs by a pair of contacting brushes by which the current is fed to what, in this case, becomes essentially the rotor of a d-c motor. A stabilizing mechanical load for the rotor is provided by means of an aluminum or copper disk carried by the shaft and rotating between the poles of one or more permanent magnets. A diagrammatic sketch of such an electrical layout is to be seen in Fig. 262. The rotor is mechanically articulated with, and drives, a system of gears which move pointers over scales whose total reading gives the number of watt-hours of electrical energy that have passed through the meter while in operation. Such an instrument will, of course, function only on direct current. And since direct current constitutes a small percentage of the total energy utilized the electrodynamometer type of instrument finds only limited application.

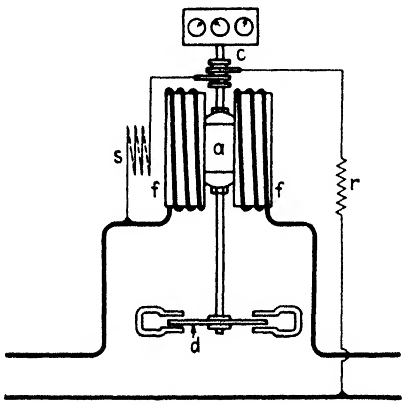
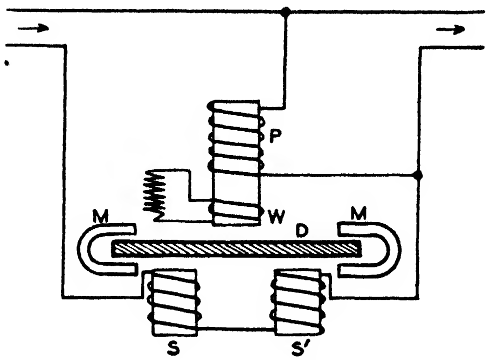


FIG. 262.—Commutator type of watt-hour meter.

The **induction type of watt-hour meter** is, essentially, a single-phase induction motor. It is provided with a magnetic load, as in the type just described. The electrical essentials of this type of unit are sketched in Fig. 263*a*. The potential coil is shown as P and the current coils as S and S'. W is a compensating winding. The rotor consists of an aluminum or copper disk D in the region of which a rotating magnetic field is set up by the quadrature currents in the potential and field windings, the former winding being highly inductive. The compensating winding W is essentially a secondary transformer winding about the core which carries the potential winding. This winding is shorted through a definite load resistance. The current induced in this circuit sets up a magnetic field, which compensates for the fact that the fields due to the potential and current coils are not strictly in quadrature, particularly if the power factor is not close to unity. The metallic disk rotates between the poles of one or more permanent magnets, thus providing a mechanical load for the motor action of the rotor. As in the electrodynamometer type, the rotor shaft causes the rotation of a series of geared indicators. An illustration of an instrument of this type appears as Fig. 263*b*. A detailed description of the induction watt-hour meter may be found in the "Standard Handbook for Electrical Engineers," 4th ed., Sec. 3—202.



(*a*)

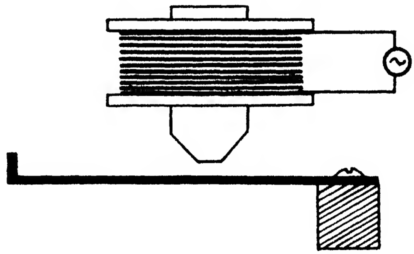
(*b*)

Fig. 263.—Induction type of watt-hour meter. (*a*) Relation of the several essential components; (*b*) Commercial model of induction type watt-hour meter. (*Courtesy of Westinghouse Electric Corp.*)

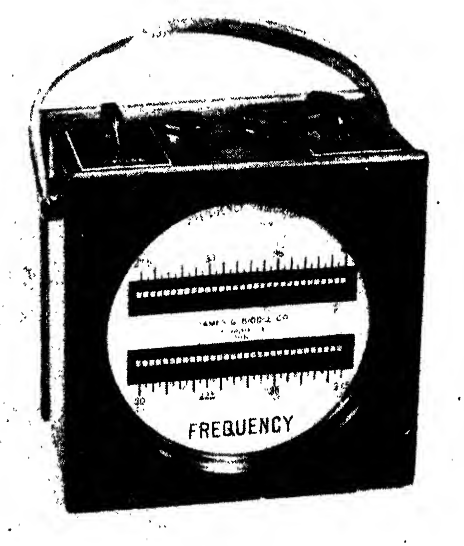
177. Frequency Indicator. The simplest and most widely used type of frequency indicator is based on the principle of mechanical resonance; it is known as the vibrating-reed instrument, though it is sometimes referred to as the **Frahm frequency meter.**

This unit consists of a series of tuned steel strips rigidly fastened at one end and free to vibrate at the other. These reeds are positioned in front of an electromagnet which is connected to the a-c electrical supply whose frequency it is desired to determine (Fig. 264*a*). The reed whose natural mechanical frequency is that of the harmonic magnetic field will be set into marked vibration. The free end of each reed is bent and painted white; and above each indicator thus formed is a number indicating the

frequency of that particular reed. When the meter is in operation, that reed whose mechanical frequency is nearest to the frequency of the a-c supply will show the greatest amplitude of vibration. Except for one or two adjacent reeds the remainder of the vibrators will show no movement.



(a)



(b)

FIG. 264.—Vibrating-reed type of frequency meter. (a) Essential components; (b) commercial model of this type of instrument. (*Courtesy of James G. Biddle Co.*)

In the actual instrument a small electromagnet drives a single vibratile member (the "armature"), which in turn is mechanically coupled to the reed assembly. By this means a given reed is thrown into vibration as a result of mechanical resonance. This type of meter is mechanically and electrically robust and can be read from a considerable distance (Fig. 264*b*).

178. Duddell Oscillograph. There are two types of oscillographs now in common use: (1) the cathode-ray assembly (Sec. 198), and (2) the moving-coil unit originally developed by Duddell. By means of an oscillograph one is able to observe visually, or photographically record, current and potential wave forms, as well as the phase relations between these factors.

The Duddell instrument is, essentially, a special form of D'Arsonval galvanometer. The moving element consists of a **single** loop of phosphorbronze or silver ribbon to which is attached a tiny mirror. This single-turn coil is stretched over grooves in small blocks of insulating material, and held taut by means of a small pulley attached to a spring, as shown in Fig. 265. The pole pieces of a strong permanent magnet produce a field at right angles to the direction of the delicate conductors. If and when current flows through the ribbon each side of the loop will be slightly deflected in a direction opposite to that of the other and at right angles to the direction of the magnetic flux. Thus the tiny mirror will be slightly rotated and a beam of light incident on the mirror will be deflected through an angle twice as great. A parallel beam of light from a straight-filament incandescent lamp is caused to fall on the movable mirror. The reflected beam is projected on to a rotating mirror system, after which,

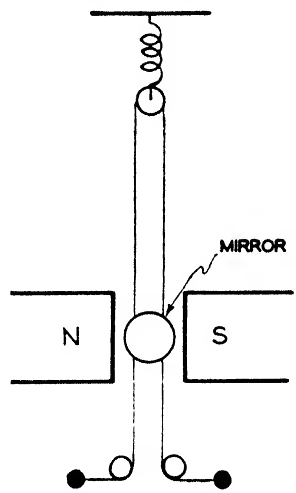


FIG. 265.—Essentials of the Duddell type of oscillograph.

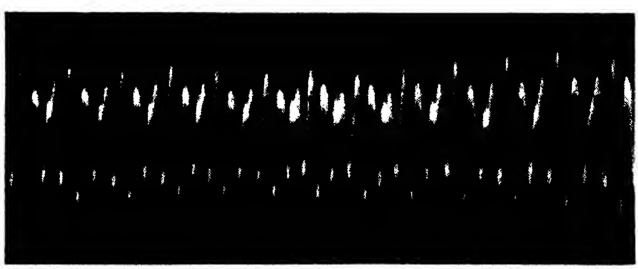


FIG. 266.—Record made by means of a Duddell oscillograph.

by means of a suitable optical system, it is brought to a focus on a ground-glass observing screen or on a photographic film. By arranging the axis of the rotating mirror at right angles to the direction of motion of the mirror, an amplitude and a time axis are provided. The reproduction of a photographic recording of a complex wave form made by an instrument of this type is reproduced as Fig. 266. The tension of the moving system

is so adjusted that its natural period is well above the frequency of the current or voltage being studied, and may be as high as 10,000 cycles/sec. Damping is effected by immersing the moving system in a light transparent oil.

In electrical engineering work it is often necessary to make records of both current and voltage simultaneously, together with time markings. In such a case a multiple-element oscillograph is employed. Such an assembly consists of several single-loop galvanometers, of the type above described, with the optical system so arranged that the wave forms appear

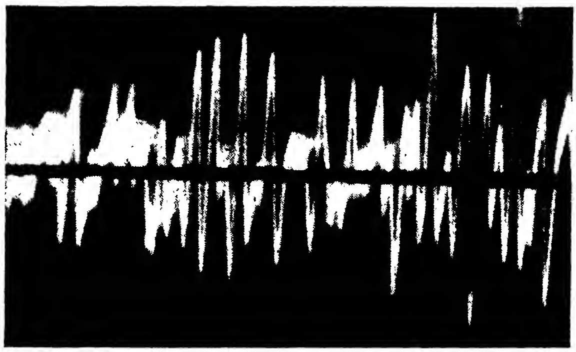


FIG. 267.—Oscillographic record of an electrical transient.

on the viewing screen simultaneously. Six or more galvanometers are sometimes employed. Since the delicate conductor constituting the single-turn loop has a very low resistance, it is necessary to limit the current that passes through the moving system to something of the order of one-tenth of an ampere. For the observation of currents this is accomplished by means of shunts and, in the case of voltage studies, by series resistors. The multiple-element Duddell instrument is particularly useful in the study of the electrical transients that occur in transmission lines when switching takes place, and also in connection with the investigation of phenomena caused by lightning and other extraneous disturbances. It is also employed in connection with geophysical prospecting. Figure 267 shows the record of a transient phenomenon.

CHAPTER XXIII

ELECTRICAL UNITS

179. Fundamental and Derived Units. Having considered the concepts and laws relating to electrostatics, magnetostatics, and electrodynamics we are now in a position to review the subject of electrical units in its broader aspects. In order to arrive at a comprehensive understanding of the interrelations which exist in any system of units it is important to note that in making a statement concerning the magnitude of any physical quantity two factors are involved, one of which is **a mere number** and the other **the unit in which the quantity is expressed.** It is this second factor which we now proceed to examine.

Obviously there are many types of quantities and hence many possible forms of units. For instance, there are mechanical, thermal, electrical, and other types of quantities. It is conceivable that we might have entirely independent systems of units in which to express the various quantities occurring in these several fields. But since experience shows that these various fields are intimately related, and in fact interdependent, it is obviously desirable to adopt some basic system of units to which all special units may be referred. Experience has also shown that all quantities with which science is now familiar can be expressed ultimately in terms of **three fundamental units,** viz., length, mass, and time; length being expressed in centimeters, mass in grams, and time in seconds, thus giving us what is known as the cgs system. In recent years there has been some discussion concerning the desirability of using the meter and the kilogram in place of the centimeter and the gram as basic units of length and mass. This so-called mks system is touched upon in the closing paragraph of this chapter.

In the electrical field there are two other physical concepts that are sometimes involved in the expression of quantities, viz., the dielectric constant K and magnetic permeability μ. However, these two latter concepts are not commonly referred to as fundamental units. Indeed some writers prefer to omit them from a discussion such as this. For reasons which will appear as we proceed, it is the author's feeling that both K and μ should enter into our discussion.

Units which are **derived** from the fundamental units just mentioned are referred to as **absolute, or cgs, units.** For instance, linear velocity is a ratio of linear displacement to time, or $v = d/t$. Since linear dis-

placement is naturally expressed in terms of length, velocity is expressed in terms of length and time. Therefore the absolute unit of velocity is the centimeter per second.

Again, it has been shown (Sec. 57) that capacitance is expressed in terms of a unit of length; hence the absolute, or cgs, unit of capacitance is in reality the centimeter, which has been named the abfarad. Other examples might be cited, but the foregoing will suffice to illustrate what is meant by absolute, or cgs units.

180. Systems of Electrical Units. At various points in the preceding chapters we have referred to three kinds of electrical units, the electrostatic, the electromagnetic, and the so-called practical or engineering units. In all electrical measurements the magnitudes with which we deal can only be determined by means of accurate observations of mechanical effects. As pointed out in our discussions, it has been found convenient to adopt the mechanical effects due to unit charge or to unit pole as a basis for electrical calculations. The magnitudes of these effects can be, and are, expressed in absolute units; and all of the basic equations which we have developed have involved only fundamental units, or units derived directly from such units. It therefore follows that both **the electrostatic and electromagnetic quantities resulting from the application of any of our basic equations are expressed in absolute or cgs units.**

It will be evident that the establishment of a common absolute system of units in the fields of mechanics, electricity, and magnetism tends to simplify and unify both the theoretical and practical aspects of these intimately related sciences. The fact that we have two systems of absolute units in electrical and magnetic computations does not present a serious difficulty, for the reason that the units in one system may be converted readily into those of the other, or into the practical units. From the foregoing statements it will be obvious why the cgs units have come to bear such names as abcoulombs, abvolts, abamperes, etc.

While the development of a system of units resulting from theoretical analysis based on the three fundamental units has done much to further the progress of electrical science, these cgs units are in some instances inconveniently large and, in other cases, too small to be applicable in practice. For instance, the absolute unit of resistance in the em system is equal to the resistance of a piece of copper wire one millimeter in diameter and about one twenty-thousandth of a millimeter in length. By international agreement (1908), as already mentioned, units of a more convenient size have been adopted for use in the electrical industry, these practical units being defined **in terms of cgs electromagnetic units.** These subsidiary units are derived from the absolute units by multiplying or dividing by powers of 10.

181. Comparison of Practical and Cgs Units. In the preceding chapters we have indicated the accepted definitions of the several practical or engineering electrical units; the cgs equivalents have also been given. For purposes of review, and for convenient reference, there follows a fairly complete table of these units and their conversion ratios.

Quantity	Practical unit	Equivalent in cgs units: Electrostatic	Equivalent in cgs units: Electromagnetic
Resistance.	Ohm	$\frac{1}{9 \times 10^{11}}$ statohm	10^{9} abohms
	Megohm	$\frac{1}{9 \times 10^{17}}$ statohms	10^{15} abohms
Emf and p.d.	Volt	$\frac{1}{300}$ statvolt	10^{8} abvolts
Current.	Ampere	3×10^{9} statamperes	10^{-1} abampere
	Milliampere	3×10^{6} statamperes	10^{-5} abampere
Quantity	Coulomb	3×10^{9} statcoulombs	10^{-1} abcoulomb
Capacitance	Farad	9×10^{11} statfarads	10^{-9} abfarad
	Microfarad	9×10^{5} statfarads	10^{-15} abfarad
Inductance	Henry	$\frac{1}{9 \times 10^{11}}$ stathenry	10^{9} abhenrys
	Millihenry	$\frac{1}{9 \times 10^{7}}$ stathenry	10^{6} abhenrys
Energy.	Joule $= 10^{7}$ ergs Watt-hour $=$ 3,600 amp-sec Kwhr $= 3.6 \times 10^{6}$ amp-sec $= 3.6 \times 10^{6}$ joules		
Power.	Watt $= 10^{7}$ ergs/sec Kilowatt $=$ 1,000 watts $= 10^{10}$ ergs/sec $=$ 1.341 hp		

182. Dimensions of Electrical Units. In dealing with physical quantities, particularly in the electrical field, it is useful to know to what extent the fundamental units of length, mass, and time enter into the unit in which a given quantity is expressed.

We know that linear velocity is the ratio of linear displacement to time; in other words, length enters once into a computation of velocity and time also enters once, but **inversely.** These facts may be expressed by what is known as a dimensional equation, thus,

$$[v] = \left[\frac{L}{T}\right] = [LT^{-1}],$$

which is to be read, "the dimensions of the unit of velocity **are** LT^{-1}," the term **dimension** meaning the extent to which the fundamental units are involved in the quantity being examined. The **exponents** of the symbols for the fundamental units indicate the **dimensions** of derived units in terms of the fundamental units. The brackets indicate that the enclosed letters refer to the **kind** of unit and not to its magnitude.

To extend our example somewhat, we might find the dimensions of acceleration, the defining equation of which is

$$a = \frac{v}{t}.$$

Since we already know the dimensions of velocity we may at once write

$$[a] = [LT^{-2}].$$

Again, force is defined thus,

$$F = ma;$$

hence the dimensional equation for force becomes

$$[F] = [MLT^{-2}].$$

It is thus seen that the fundamental unit of mass enters once in the unit of force, the unit of length also once, and the unit of time inversely twice.

The dimensions of the electrical units are built up in a similar manner. Beginning with the mechanical force action between charges, the dimensions of the absolute (cgs) electrostatic units are readily determined. The force between two charges is given by Eq. (2) in the form

$$F = \frac{qq'}{Kd^2}.$$

Making $q = q'$ we have

$$F = \frac{q^2}{Kd^2};$$

or

$$q = d\sqrt{FK} = dF^{1/2}K^{1/2}.$$

The dimensional equation for electrical quantity then becomes

$$[q] = [L(MLT^{-2})^{1/2}K^{1/2}] = [M^{1/2}L^{3/2}T^{-1}K^{1/2}].$$

By continuing this process the dimensions of all of the electrostatic units may be indicated.

The dimensions of the cgs electromagnetic units can be set up in a similar manner. Starting with the force between two magnetic poles as

given by Eq. (16) we have

$$F = \frac{mm'}{\mu d^2}$$

where μ is the magnetic permeability. If we make the two poles of equal value the above equation becomes

$$F = \frac{m^2}{\mu d^2}.$$

Solving for m, as we did in the corresponding electrostatic equation, we get

$$m = d\sqrt{\mu F} = dF^{1/2}\mu^{1/2}$$

from which the dimensional equation for the unit of pole strength may be written thus,

$$[m] = [M^{1/2}L^{3/2}T^{-1}\mu^{1/2}].$$

This line of reasoning may be extended to other quantities in the electromagnetic system. Take, for instance, the case involving magnetic field intensity as expressed in Eq. (18), which is

$$F = mH,$$

or

$$H = \frac{F}{m}.$$

On the basis of this defining equation the dimensional equation would be

$$[H] = [M^{1/2}L^{-1/2}T^{-1}\mu^{-1/2}].$$

Thus we have a statement of the extent to which the fundamental units of length, mass, and time enter into the cgs unit of magnetic field intensity, which is the oersted.

One further example of the method by which the dimensions of electrical units are arrived at will suffice. The relation between the field strength due to the current in a long straight conductor and the magnitude of that current is given by Eq. (141) as

$$H = \frac{2I}{x},$$

or

$$I = \frac{xH}{2}.$$

The factor 2 may be neglected in setting up the dimensional equation,

$$[I] = [LM^{1/2}L^{-1/2}T^{-1}\mu^{-1/2}] = [M^{1/2}L^{1/2}T^{-1}\mu^{-1/2}],$$

thus giving the dimensions of the absolute emu of current. By a continuation of this general process the dimensions of the remaining emu may be deduced. Following is a tabular list of the dimensions of a number of units in both the electrostatic and electromagnetic systems.

Unit of	Dimensions in		Ratio of electrostatic to electromagnetic units
	Electrostatic system in terms of L, M, T, K	Electromagnetic system in terms of L, M, T, μ	
Quantity	$L^{3/2}M^{1/2}T^{-1}K^{1/2}$	$L^{1/2}M^{1/2}\mu^{-1/2}$	v
Electrostatic field intensity	$L^{-1/2}M^{1/2}T^{-1}K^{-1/2}$	$L^{1/2}M^{1/2}T^{-2}\mu^{1/2}$	v
Potential	$L^{1/2}M^{1/2}T^{-1}K^{-1/2}$	$L^{3/2}M^{1/2}T^{-2}\mu^{1/2}$	v^{-1}
Capacitance	LK	$L^{-1}T^{2}\mu^{-1}$	v^{2}
Resistance	$L^{-1}TK^{-1}$	$LT^{-1}\mu$	v^{-2}
Current	$L^{3/2}M^{1/2}T^{-2}K^{1/2}$	$L^{1/2}M^{1/2}T^{-1}\mu^{-1/2}$	v
Inductance	$L^{-1}T^{2}K^{-1}$	$L\mu$	v^{-2}
Pole strength	$L^{1/2}M^{1/2}K^{1/2}$	$L^{3/2}M^{1/2}T^{-1}\mu^{1/2}$	v^{-1}
Magnetic field intensity	$L^{1/2}M^{1/2}T^{-2}K^{1/2}$	$L^{-1/2}M^{1/2}T^{-1}\mu^{-1/2}$	v
Magnetic flux	$L^{1/2}M^{1/2}K^{-1/2}$	$L^{3/2}M^{1/2}T^{-1}\mu^{1/2}$	v^{-1}

Dimensional equations serve a number of useful purposes. One important utility is that they may be made to serve as a check on the logic by which any given working equation has been derived, and thus upon its correctness. In any equation all of the terms must have the same dimensions. As an example, suppose one were to consider the equation representing Ohm's law $I = E/R$. The dimensional equation, using em units, would be

$$L^{1/2}M^{1/2}T^{-1}\mu^{-1/2} = \frac{L^{3/2}M^{1/2}T^{-2}\mu^{1/2}}{LT^{1}\mu} = L^{1/2}M^{1/2}T^{-1}\mu^{-1/2},$$

thus showing that the equation is correctly stated. If the two sides of the dimensional equation had turned out to be unlike, it would have indicated that there was something incorrect in the basic relation.

In addition to the utility of dimensional formulas in checking the correctness of equations by which calculations are to be made, a knowledge of the dimensions of the units entering into a given relation will often assist one in arriving at a conclusion as to the nature of the units in which the resulting quantity must be expressed. This knowledge will also serve to show how the fundamental units enter into the final result.

A case in point is that of inductance. The dimension of inductance in emu is length to the first power. The cgs unit is therefore the centimeter.

Furthermore, dimensional formulas sometimes serve to reveal intimate relations between magnitudes which might, by simple inspection, appear to be quite independent. This utility of dimensional formulas is strikingly shown by an examination of the ratio of the dimensions of the various electrical units as given in the two systems of cgs units. Take, for instance, the case of the unit of current. Assume the conductor to be in free space, thus making both K and μ unity. The ratio of the electrostatic to the electromagnetic dimensions is

$$\frac{[L^{3/2}M^{1/2}T^{-2}]}{[L^{1/2}M^{1/2}T^{-1}]} = [LT^{-1}] = [v].$$

It is thus evident that the ratio of the cgs electrostatic and electromagnetic units of current is of the nature of velocity.

Again, if we examine the ratio of the dimensions in the case of the units of potential we find that

$$\frac{[L^{1/2}M^{1/2}T^{-1}]}{[L^{3/2}M^{1/2}T^{-2}]} = \frac{1}{[LT^{-1}]} = \left[\frac{1}{v}\right];$$

and for capacitance the ratio is

$$\frac{[L]}{[L^{-1}T^{-2}]} = [L^2T^{-2}] = [(LT^{-1})^2] = [v^2].$$

If we were to continue this process for the remaining units we should have the ratio values shown in the last column of the table on page 394. An examination of the ratios there listed discloses a most remarkable fact; namely, that **in all cases the ratio is some power of a velocity or a reciprocal of a power of a velocity.** The question then, naturally, presents itself as to the numerical value of the velocity term. Fortunately it is not difficult to make at least an approximate determination of this factor. It is possible to construct accurately a condenser and to compute its capacitance from its geometrical dimensions. The result will be in cgs electrostatic units. By charging the same condenser to a known potential and discharging it through a ballistic galvanometer we may measure its capacitance in cgs electromagnetic units. If, then, we take the ratio of the two capacitance values thus obtained, we find it to be approximately $(3 \times 10^{10})^2$. We are thus confronted with the astonishing fact that the ratio of the two absolute systems of electrical units involves a velocity, and that this velocity, as closely as can be determined experimentally, **is the same as the velocity of light in free space.** This significant relation was one of the factors that led Maxwell to advance the

theory **that light is essentially an electromagnetic phenomenon** (Sec. 242).

In passing, one other aspect of the interpretation of dimensional formulas should be noted. Many of the dimensions appear as fractions. It is difficult to attach any meaning to such an expression as $L^{3/2}$ or $M^{1/2}$ in and by themselves. Permeability μ and the dielectric constant K are factors which serve to specify, in certain limited respects, the properties of the medium. We do not know the absolute dimensions of K and μ. If we had this information, and it was incorporated in our dimensional equations, it is possible that those terms which now appear with fractional exponents would be rationalized, and it might be that the dimensions of the units in the two systems would prove to be the same. In other words, we do not at present know what the mechanism is by which one charge attracts another or by which one pole exerts a force upon another. It may, however, prove to be significant in this connection that the product of K and μ may be expressed in terms of velocity. This relation between K, μ, and v may be deduced readily by equating the es and em dimensional expressions for any one of the units, say capacitance. The writing of such an equality is justified because of the fact that it is unreasonable to assume that any given quantity can have essentially two different sets of dimensions of the fundamental units. Proceeding on this basis we have

$$[LK] = [L^{-1}T^2\mu^{-1}],$$

which yields

$$\left[\frac{L^2}{T^2}\right] = \left[\frac{1}{K\mu}\right] = [v^2],$$

or

$$[v] = \left[\frac{1}{\sqrt{K\mu}}\right].$$

This relation indicates that $1/\sqrt{K\mu}$ has the dimension of a velocity. When we have more complete information concerning the properties of free space it is possible that we may be able to assign definite dimensions to the dielectric constant and to permeability.

In any event the fact that $1/\sqrt{K\mu}$ has the dimension of a velocity, together with the fact previously mentioned, served as a basis for Maxwell's theory of electromagnetic radiation—a theory which has been strikingly confirmed since Maxwell's time, and which will be considered in the final chapter of this book.

183. Definition of Electrical Units Adopted by the AIEE. The so-called international system of electrical units, originally adopted in 1893 and modified in 1908, was founded on the four fundamental quantities of **resistance, current, length,** and **time.** In connection with the

discussion of resistance (Sec. 70) and current (Sec. 67) the basis of the international definitions of the units of resistance and current was indicated. Progress in the refinement of electrical measurements showed that slight numerical discrepancies existed between these international units and the units as defined by the fundamental equations (absolute units). In 1935 the International Commission of Weights and Measures decided to discontinue the use of the international units. In the future **the absolute units are to be the cgs electromagnetic units,** and the practical units are to be certain multiples or fractions of the absolute units. On Aug. 12, 1941, under the sponsorship of the American Institute of Electrical Engineers, the American Standards Association approved a set of definitions of electrical terms. These definitions were approved by the Canadian Engineering Standards Association on Mar. 2, 1942. As world conditions again become settled, other leading countries will probably subscribe to the definitions thus established. Below we give excerpts from the above-mentioned statements regarding several of the more common units.

The cgs electromagnetic unit of current, or abampere, is defined by means of the law connecting the current in an electric circuit with the magnetic intensity at any point in its magnetic field.

The law is given by the equation

$$H = kI \oint \frac{[\bar{r} \times d\bar{s}]}{r^3}.$$

In the above equation

H is the magnetic intensity in oersteds at any point P in the magnetic field of a circuit carrying a current I; k is a proportionality factor which is given the value of unity to establish the cgs unit of current; r is the magnitude of the vector $\bar{r}$, representing the distance in centimeters from the point P to an element $d\bar{s}$ of the electric circuit; and the line integral of the vector product of $\bar{r}$ and $d\bar{s}$ is taken completely around the electric circuit.

The ampere is one-tenth of the abampere.

The cgs electromagnetic unit of emf is the emf in a circuit when, with one abampere of current flowing, energy is converted into other kinds of energy at the rate of one erg per second.

The volt is equal to 10^8 abvolts.

The cgs unit of resistance is the resistance of a conductor when, with an unvarying current of one abampere flowing through it, the potential difference between the ends of the conductor is one abvolt.

An ohm is equal to 10^9 abohms.

The practical units defined above are referred to as the **absolute** practical or engineering units.

184. Determination of the Absolute Values of the Ohm and the Ampere. From the definitions just given, and from others that might be listed, it is evident that it becomes necessary to determine, **by experimental means,** the numerical values of but two quantities, viz., the ohm and the ampere. Various procedures are followed for carrying out these determinations.

In determining the **absolute value of the ohm** a method suggested by Lorenz is often followed, and may be taken as representative. The **Lorenz method** consists in balancing the drop across the resistance being

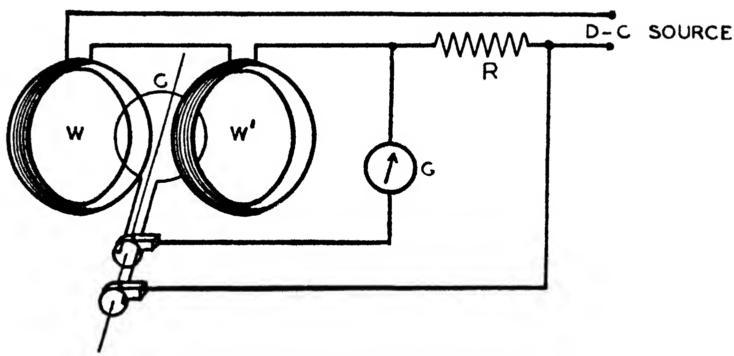


Fig. 268.—Determination of the absolute value of the ohm.

studied against the emf generated in a coil being rotated in the magnetic field developed by a pair of Helmholtz coils. The circuit setup is sketched in Fig. 268. W and W' are the Helmholtz coils, C the rotating coil, R the resistance, and G a galvanometer. By Eq. (145) the field at the center of the coils will be

$$H = \frac{2.86\pi}{r} I = BI,$$

where B is the constant of the coils. From Eq. (178) it will be seen that the peak value of the emf developed in the rotating coil C will be given by the expression

$$E_{\max} = 2\pi(RPS)NAH,$$

where A is the area of the coil and N the number of turns in the rotor winding. Therefore, $E_{\max} = 2\pi(RPS)NABI$. When the angular speed of the rotor is so adjusted that the peak emf developed therein equals the d-c drop over the resistor R the galvanometer will show zero deflection. When this condition obtains

$$2\pi(RPS)NABI = RI,$$

which leads to

$$R = 2\pi(RPS)NAB. \tag{242}$$

Thus we see that, aside from the numerical constants of the equation, the only factors involved are **length** and **time.** The absolute value of the unit of resistance may therefore be expressed in these two fundamental units. The accuracy with which such a determination can be carried out will depend upon how accurately one is able to measure the dimensions of the several coils and the speed of revolution. Modern stroboscopic methods make possible a high degree of accuracy in the determination of angular speed.

In the determination of the **absolute value of the ampere** mechanical force serves as the basis of the measurement. Referring to the table on

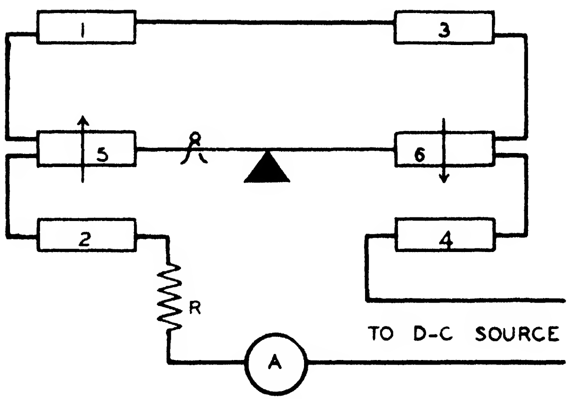


FIG. 269.—Kelvin balance.

page 394 it will be seen that the dimensions of current in the em system are $L^{1/2}M^{1/2}T^{-1}\mu^{-1/2}$ or $\sqrt{LM}/T\mu$. It will be remembered that the expression ML/T^2 gives the dimensions of **force.** If, then, we omit, for the moment, any consideration of the permeability factor, we see that the remaining dimensions of current are those of the square root of force. Hence it should be possible to determine the magnitude of a current **in terms of force.** Since it is not feasible to make such a determination *in vacuo* it will be necessary to correct the experimental value found by dividing by the square root of μ for air. Obviously this correction is very small and, for any but the most refined measurements, may be neglected.

Lord Kelvin developed a current balance by means of which it is possible to carry out experimentally the determination indicated by the foregoing analytical discussion. A rough diagrammatic sketch of the **Kelvin balance** is to be seen in Fig. 269. Omitting electrical and mechanical details, it may be said that the assembly consists essentially of six

similar coils connected in series. Four of these coils (1, 2, 3, and 4) are fixed in position while the remaining two (5 and 6) are supported by a mechanical member that is pivoted at its center. By suitable means, current is led into and out of the movable coils without mechanically disturbing the system. Coils 5 and 6 are in the fields produced by the coil pairs 1–2 and 3–4. The fields due to the current in 5 and 6 react with the fields in the region of the fixed pairs in such a manner that oppositely directed forces are established, as indicated by the arrows. Thus the beam and its attached coils will rotate about the central axis. Provision is made for restoring the balance by means of a weight which may be slid along a scale (in length units) attached to the beam. It may be shown that the current through the coils will be given by the relation

$$I = a\sqrt{mg}, \tag{243}$$

where the effective force mg is determined by the position of the weight on the beam when equilibrium is restored, much as in the use of a rider when weighing with a chemical balance. a is a constant depending for its numerical value on the dimensions of the coils, their relative positions, and the number of turns in each. Thus it is possible to determine the magnitude of a current in terms of a force; and a current-indicating instrument (A in the sketch) can thereby be calibrated. If desired, the drop over a standard resistance R can be balanced against the emf of a standard cell by means of a potentiometer; and thus the standard cell can be calibrated. In either event, the magnitude of the current is determined **in terms of force units.**

In their most refined form both the Lorenz assembly and the Kelvin balance are complicated and delicate organizations. They are therefore to be found only in well-equipped standardizing laboratories. Most educational and commercial laboratories rely upon well-calibrated resistances and standard cells as a basis for routine measurements. In this country the Bureau of Standards and, in England, the National Physical Laboratory, serve as the ultimate check on the calibration of all electrical-measuring instruments and devices.

In concluding our discussion of units reference should, perhaps, be made to a general system of units originally suggested in 1901 by the Italian physicist, Giorgi. The Giorgi plan involves an absolute system of units which is based on the meter, kilogram, and second, instead of the centimeter, gram, and second; it is accordingly referred to as the **mks system.** In certain respects such a system has several advantages, but when extended to cover electrical units serious disadvantages appear. Neither time nor space permit a review of the arguments for and against such a system. It must suffice to point out that a new unit of force (the

newton) is involved. The familiar second law of motion becomes

$$F \text{ (newtons)} = m \text{ (kilograms) } a \text{ (meter/sec}^2\text{)}.$$

Carried over into the electrical field we find that the permeability of a vacuum turns out to have a numerical value of 10^{-7}, and the dielectric constant has a value of $\frac{1}{9} \times 10^9$. Furthermore, in the mks system, both μ and K have dimensions. The International Electrochemical Commission in 1935 approved the mks system and set Jan. 1, 1940 as the date when the system should go into universal use. Quite aside from the effect of the war on such a matter, there is as yet no agreement among electrical engineers as to how the proposed system is to be applied to magnetic and electrostatic units. We shall, accordingly, follow the more familiar system of electrical units.

CHAPTER XXIV

ELECTRONIC EMISSION

185. Types of Emission. In our discussion thus far we have been dealing with electrons which were intimately associated with solids or liquids. In the case of solids we have thought of the orderly migration of so-called **free** electrons under the influence of electric or magnetic forces; and it was agreed that this directed electronic movement constitutes what we know as the electric current. We are now to consider an important group of phenomena that involves the motion of electrons which have been temporarily, but completely, **detached from the atomic structures of any solid, liquid, or gas.** In other words, we are to deal with **electrons in free space.** In any such discussion the question which naturally first presents itself is: How is a supply of free electrons to be made available?

There are four principal means whereby electrons may be completely detached from the atoms constituting a solid, and two of these same methods may also be utilized to liberate electrons from the atoms of a gas. These methods are

1. The application of radiant energy, giving rise to **photoelectric emission.**
2. The application of heat, giving rise to **thermionic emission.**
3. The application of an intense electrostatic field in the region of a solid, thus giving rise to **field emission.**
4. Bombardment of a surface by electric charges moving at high velocities, thus giving rise to what is called **secondary emission.** As we shall see, the first and last methods may be applied in the case of gases.

The liberation of an electron from an atomic structure involves the transfer of energy to the electron from an outside source. Work must be done in overcoming the surface constraint due to the attraction of a nucleus. The amount of energy involved in such a process of liberation will depend upon the nature of the atom from which the electron is being separated. The energy which must be transferred to an electron to cause it to cross the bounding surface of a metal is known as the **work function** of that particular substance. The work function of many materials has been determined experimentally and found to be of the order of 10^{-12} to

10^{-11} ergs per electron—a small but very definite amount of energy. The symbol ϕ_0 is commonly employed to designate the work function.

In the study of electronics it has been found convenient to express the work function in terms of the potential difference through which an electron must be taken in order to give it the energy equivalent of the work function. By letting V represent $V_2 - V_1$ in Eq. (53), putting e (the electronic charge) for q, and transposing we have

$$V = \frac{W}{e},$$

where W is the energy required to move an electron through a potential difference of V. Substituting in this equation the highest value for the work function, as given above, we get

$$\begin{aligned} V &= \frac{10^{-11}}{4.8 \times 10^{10}} = \frac{1}{48} \text{ statvolt} \\ &= {}^{300}\!/_{48} = 6.25 \text{ volts.} \end{aligned}$$

It is therefore evident that, under the conditions cited, an electron must move through a potential difference of 6.25 volts in order to acquire 10^{-11} erg of energy. Energy and potential difference thus being so intimately related ($W = eV$), it is convenient to speak of the energy imparted to an electron in terms of the potential difference involved. Accordingly, **the electron-volt may be defined as the change in energy experienced by an electron when it is taken through a potential difference of one volt. Its numerical value is approximately 1.591 $\times$ 10^{-12} erg.** Below is a list of the work-function values of a number of substances which are frequently used as a source of electrons.

Substance	Work function ϕ_0, in electron-volts
Platinum (Pt)	6.26
Tungsten (W)	4.52
Thoriated tungsten (W + Th)	2.63
Barium-strontium oxides (BaO + SrO)	1.04
Magnesium (Mg)	2.42
Potassium (K)	2.24
Sodium (Na)	2.46
Caesium (Cs) (thin film on silver)	1.81

186. Photoelectric Emission. Electrons may be set free from their parent atomic structures by means of radiant energy. This phenomenon was first observed in 1887 by Hertz, of whom we shall hear later. He noted that the electrical discharge between two terminals was facilitated when the negative electrode was illuminated and that the effect was most

pronounced when the illumination consisted of wave lengths in the ultra-violet region. Following up the original observations by Hertz, Hallwachs, using a clean zinc surface, carried out a series of fundamental experiments which led to the conclusion that negative electricity may be liberated from a metallic surface through the agency of short-wave radiant energy.

Beginning in 1889, Elster and Geitel began an extended series of investigations involving this phenomenon. They found that certain elements, particularly the alkali metals such as sodium and potassium, yielded a more copious supply of electrons than did most other substances. These investigators discovered that the photoelectric effect could be enhanced by inclosing the emitting surface in a vacuum. They also made the fundamental observation that the **photoelectric emission was directly proportional to the light intensity.** In 1899, Lenard and J. J. Thomson showed that the electrical entities released by light are what we now know as electrons. Lenard found that **the kinetic energy of the liberated electrons is independent of the intensity of the incident radiant energy,** but that the **number** of charges emitted **is** proportional to it.

It is difficult to explain these observed facts on the basis of the wave theory of light. Einstein was, accordingly, led to suggest, in 1905, that we might apply Planck's quantum theory to the phenomenon of photoelectric emission. According to this point of view the light quanta (photons) disappear as such, their energy content being absorbed by the electrons within the metal which, in turn, escape as photoelectrons. The kinetic energy of an escaping electron should, on this basis, be equal to the energy of a quantum minus the energy required to free the electron from the emitting surface. In mathematical form this statement would be

$$h\nu - w = \tfrac{1}{2}mv^2, \tag{244}$$

where h is Planck's constant, ν the frequency of the incident radiation, w the energy required to overcome the surface restraint, and v the maximum velocity. The factor w is the photoelectric work function, and corresponds to the term which we have previously designated by ϕ_0. It will thus be seen that the energy of an escaping electron is necessarily less than the energy of the photon from which it acquired its increment of energy. Electrons that are liberated from atoms below the surface layer will have less kinetic energy than those that escape from the surface atoms. Consequently the velocity of the escaping electrons will not be the same for all, the range being from a maximum to zero. Bearing in mind that the energy magnitude of a photon depends upon the frequency ν involved, it will be evident that there is a minimum frequency for which the photon will contain just enough energy to cause a **surface** electron to

escape with zero velocity. This frequency is referred to as the **threshold frequency.** For each material there is a definite threshold frequency, or wave length, the value of which will depend upon the photoelectric work function of the substance involved.

The relation between the photoelectric work function and threshold frequency is easily determined. For the threshold frequency the kinetic energy term of Eq. (244) becomes zero; hence $h\nu_0 = W$, where ν_0 is the threshold frequency. We have seen that energy can be expressed in terms of potential; therefore we may write

$$h\nu_0 = Ve = \phi_0 e$$

or

$$\nu_0 = \frac{\phi_0 e}{h}. \tag{245}$$

If, then, we know the work function of a given material, we can readily compute the value of the lowest frequency to which that emitting surface will respond.

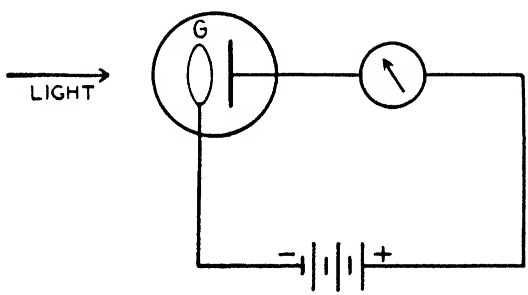


FIG. 270.—Photoelectric emission.

It is possible to determine the velocity of the fastest liberated electrons. The experimental setup to carry out such a determination is shown in Fig. 270. For a given emitting surface and a given wave length of incident radiant energy it will be found that there is a definite negative potential which, when applied to the electrode G, will prevent **any** electrons from leaving the surface. This is called the **stopping potential.** When this condition obtains it follows that

$$\tfrac{1}{2}mv^2 = Ve, \tag{246}$$

where V is the stopping potential and e the electronic charge. This follows from the facts cited above. Since the value of all the terms in the above relation are known except v, it is therefore possible to compute the magnitude of the maximum velocity. If the exciting wave length is changed, the value of the stopping potential will be changed

accordingly. A graph showing the relation between stopping potential and frequency is to be seen in Fig. 271. The relation is seen to be linear, the value ν_0 being the threshold frequency for a given surface.

In view of what has been said above it will be seen that we might combine Eqs. (244) and (246) to get

$$Ve = h\nu - w$$

or

$$V = \frac{h}{e}\nu - \frac{w}{e}. \tag{247}$$

Equation (247) applies to the graph shown in Fig. 271; the slope of the line will be given by h/e. Since we know the value of e, and since $h/e = \tan\theta$, we may readily determine the value of Planck's constant h.

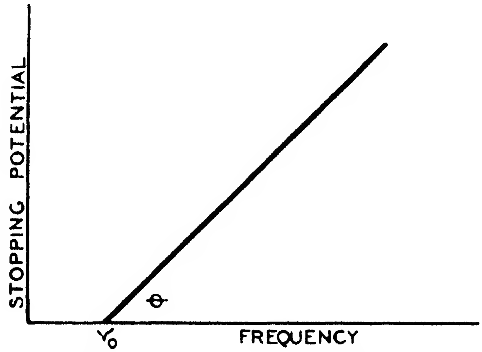


FIG. 271.—Relation between the frequency of the incident radiation and the stopping potential.

By such a procedure Millikan determined very accurately the value of Planck's constant. He found h to have a value of 6.56×10^{-27} erg-sec, which was very close to the value originally given by Planck as a result of his study of black-body radiation. It would thus appear that h is a constant of nature.

We have dwelt at some length upon the theory of photoemission because it is such a fundamental process and because the relation between quantity of radiant flux and the number of electrons liberated from a given surface is one of the most dependable laws of physics. The truth embodied in Eq. (244) (Einstein's equation) is one of the cornerstones upon which the structure of modern physics is built and was largely responsible for the awarding of the Nobel prize to its author. The determination of Planck's constant by the photoelectric method, together with his determination of the magnitude of the electronic charge, served to bring the Nobel prize award to Millikan. These two awards indicate

the theoretical significance attached to photoelectric phenomena. In the following section we shall consider the practical results of these discoveries.

187. The Photoelectric Cell. This device, as commonly made, consists of an evacuated glass enclosure in which are sealed two electrodes. The cathode usually consists of a semicylindrical piece of metal upon the concave surface of which is deposited the emitting material. A single wire or loop serves as the anode. A photograph of a typical cell is reproduced in Fig. 272. The basic circuit in which the cell ordinarily functions is shown in Fig. 273. As radiant energy falls upon the emitting surface the liberated electrons are attracted to the anode (positive terminal) and those lost by the cathode (negative terminal) are replaced by those supplied by the source of potential B. Thus an electronic current is established in the external circuit which includes the load impedance R.

FIG. 272.—Typical photoelectric cell.

The photosensitive surface which is now most commonly used consists of a silver plate, the surface of which has been carefully oxidized. Upon this silver oxide surface is deposited a layer of caesium of monatomic thickness. The thickness of the caesium layer has a bearing on the response at the shorter wave lengths, and also on the maximum response. The photosensitive layer above described is often referred to as a caesium-oxygen-silver surface.

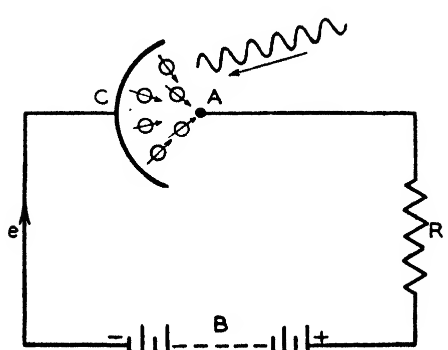


FIG. 273.—Basic photoelectric tube circuit.

There are two characteristics of a photoelectric cell which are particularly important in connection with their use in certain practical applications. These are the luminous sensitivity and the spectral response. The term **luminous sensitivity** refers to the magnitude of the electronic current available for a given amount of incident luminous flux. The values range from a fraction of a microampere to something of the order of 150 microamperes/lumen, depending on the character of the emitting surface and other structural details. Graph I of Fig. 274 shows

the response of a representative vacuum-type commercial tube. It will be observed that the response is linear. If the light flux is variable in quantity, the resulting electronic current shows practically no lag. In

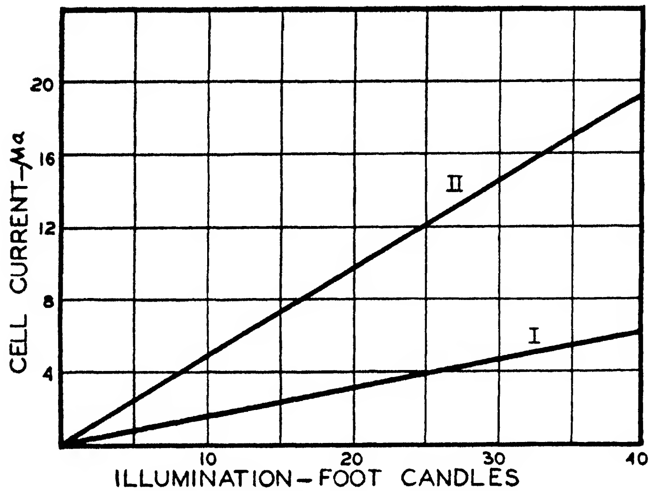


FIG. 274.—Response curves for photoelectric cells. I, vacuum type; II, tube containing a trace of gas.

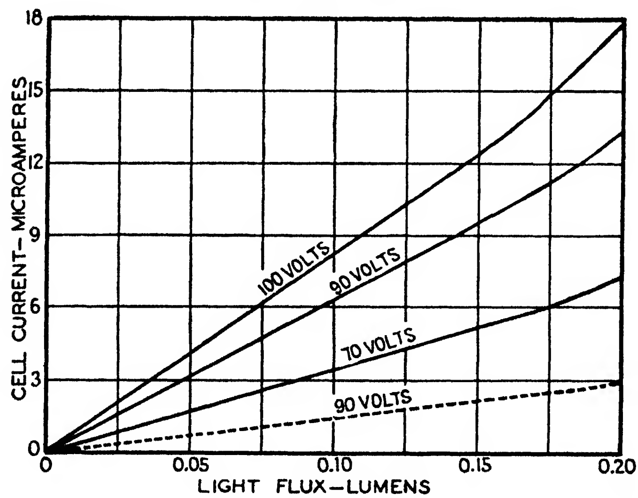


FIG. 275.—Relation between incident light flux and cell response at different anode potentials. Dotted graph is for a vacuum-type cell; others for gas-filled cells.

other words, such cells show a nearly uniform frequency response between, say, 20 and 10,000 cycles.

The magnitude of the photoelectric current is a function not only

of the light intensity, but also of the potential difference which is applied between the anode and the cathode. For the vacuum-type cell this relation is linear, as indicated by the dotted line in Fig. 275. Eighty to ninety volts is the potential difference commonly employed with this type of unit.

The **spectral response** of a photoelectric cell indicates the relation between the electronic emission and the wave length (or frequency) of the incident illumination. This relation is **not** linear, as is to be seen from an examination of the wave length *vs.* response curve (Fig. 276) for a caesium-oxide-silver emitting surface. It is to be noted that such a cell responds

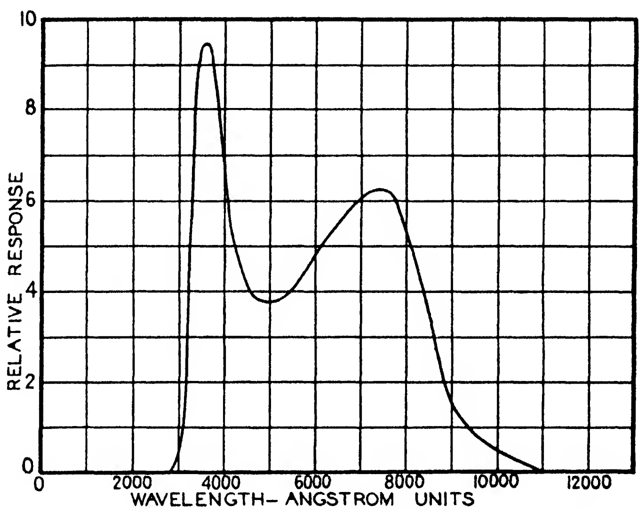


FIG. 276.—Spectral response of photoelectric cell.

strongly to wave lengths in the region of 3,500Å (blue) and also in the region of 7,500Å (deep red). The spectral response curve depends upon the nature of the material used as an emitting surface and also upon the manner in which a given surface is prepared. It is therefore possible to produce cells having special response curves.

Reference has been made above to the vacuum-type photoelectric cell. In that type of unit the current is wholly electronic. There is a type of cell into which a small amount of inert gas is admitted before it is sealed off; such a unit is called a **gas-filled cell.** The presence of an inert gas serves to augment the current obtained from the tube. The process by which this increase in current is brought about involves two other methods of liberating electrons—processes which will be discussed later and will be only briefly referred to here.

When a suitable potential difference is established between the

cathode and anode, the energy of the photoemitted electrons may become great enough to detach some electrons from the gas atoms which are present at the time. In other words, the gas becomes ionized. These additional electrons serve to augment the number originally liberated from the cathode and thus the magnitude of the electronic current is increased. But this is not all of the story. The remaining part of each atom that has lost an electron, as a result of electronic bombardment, now constitutes **a positive ion.** These positive ions will be attracted toward the cathode (negative electrode), and during their

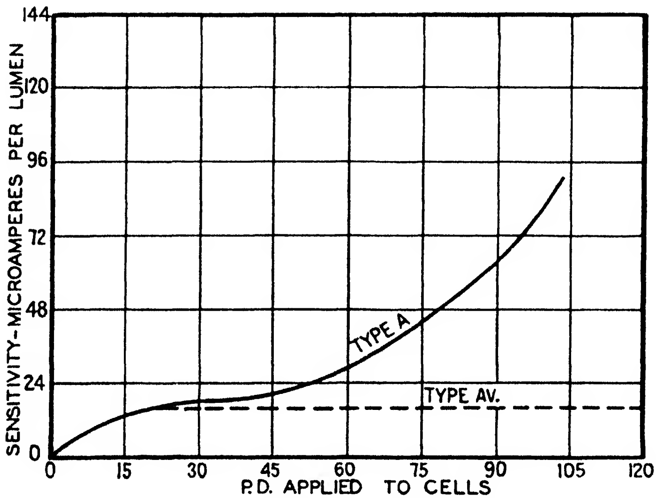


FIG. 277.—Relation between cell sensitivity and applied potential. Curve A is for a cell containing gas; curve AV for one of the vacuum type.

journey they may acquire sufficient kinetic energy to detach additional electrons **from the emitting surface** because of their impact on the cathode. Due to this secondary emission, as it is called, the supply of electrons is further augmented. Many modern photoelectric tubes are of the gas-filled type. The relation between light flux and response is not quite so linear as in the case of the vacuum-type cell, but the current output is considerably greater. The full-line curves in Fig. 275 indicate the flux-response relation in the case of typical gas-filled cells; the broken line represents the situation in the case of a vacuum-type cell.

In using gas-filled cells care must be taken not to have the cathode-anode voltage too high, or else the ionization process will become self-sustaining and the cathode surface will be damaged due to excessive ionic bombardment. The manufacturer of each gas-filled tube always states the safe potential difference that may be used in order to avoid excessive

ionization. Figure 277 gives a comparison of the effect of the magnitude of the voltage which is applied to both types of cells. These graphs should be carefully studied. (Why is the graph for the type AV cell flat beyond about 30 volts?)

The research and commercial uses to which photoelectric cells are put are varied and widespread. Photometry, talking motion pictures, telephotography, television, spectrophotometry, and light control are but a few of the important processes in which this versatile device finds essential application.

For an extended and authoritative treatment of the theoretical and experimental aspects of photoelectric phenomena the reader should consult a volume entitled "Photoelectric Phenomena" by Hughes and DuBridge.

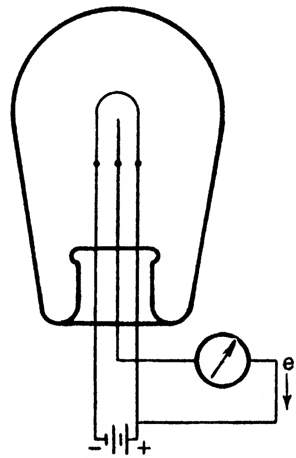


FIG. 278.—Edison effect. (*After Koller.*)

188. Thermionic Emission. Historically, the liberation of electrons by thermal means dates back to an observation made by Thomas Edison in 1883. During his work on the development of the incandescent lamp, Edison on one occasion introduced an extra electrode into one of his lamps, as shown diagrammatically in Fig. 278, and connected this terminal through a galvanometer to the **positive** terminal of the battery which supplied electrical energy to the lamp filament. When the filament was lighted, he observed that a current existed in this auxiliary circuit. However, when this special electrode was connected to the **negative** side of the battery, no current appeared. At that time we knew nothing about electrons, but what Edison had observed was an electronic current flowing from the heated filament to his special electrode and thence back to the battery. Because it was connected to the positive side of the battery, the third electrode was functioning as an anode with respect to the filament. Though Edison took out a patent covering this phenomenon, he apparently never followed up his original investigation. The phenomenon, however, bears his name and is known as the **Edison effect.** Later this effect was studied by a number of other investigators, and more particularly by Professor Fleming in England, to whose work we shall have occasion to refer later. In due time Thomson showed that the electrical entities emitted by hot bodies are what we now know as electrons. How may we account for the liberation of electrons by thermal means?

From time to time reference has been made to the fact that in a metal there are large numbers of electrons which are in constant random

motion, but that these **free** electrons are held within the body of the solid by means of nuclear forces. In discussing the subject of photoelectric emission it was pointed out that, if and when additional energy were imparted to these wandering electrons, some of the electrons might overcome the existing surface restraint and thereby escape into free space. The Edison effect is to be explained on the basis that the **thermal energy** due to the passage of the filament current **serves to increase the kinetic energy of the free electrons.** In those cases where the augmented electronic kinetic energy is equal to or exceeds the work necessary to overcome the surface restraint electronic emission will occur, *i.e.*, when $\frac{1}{2}mv^2 \geqq \phi_0$ we have what we speak of as thermionic emission. ϕ_0 is the thermionic work function and its numerical value, in the case of a given metal, is the same as that of the photoelectric work function.

In 1901, Richardson began an analytical investigation of the phenomenon of thermionic emission. Many of Professor Richardson's observations are embodied in his book entitled "The Emission of Electricity from Hot Bodies." Later Dushman and others also attacked the problem analytically, with the result that we have available an equation that gives the emission current per unit area of a heated filament in terms of the absolute temperature and certain constants. This relation, known as the Richardson-Dushman equation, has the following form:[1]

$$I_e = AT^2\epsilon^{-\frac{\phi_0}{kT}}, \tag{248}$$

where I_e is the maximum emission current in amperes per square centimeter, T the absolute (Kelvin) temperature, A a constant, which for pure metals has a value of 60.2, and k Boltzmann's constant

$$(8.63 \times 10^{-5} \text{ electron-volts/deg}).$$

It should be said that in the case of impure metals A may have a value differing greatly from that given above. The exponential term of Eq. (248) is sometimes written in the form $\epsilon^{-(b_0/T)}$, where $b_0, = \phi_0/k$. It is to be noted that Richardson's equation holds only when the anode potential is sufficiently high to attract all free electrons to the anode. An examination of the Richardson-Dushman equation discloses the fact that the exponential term has the most significance, so far as the relation between I_e and T is concerned. The currently accepted values for A and b_0, in the case of several cathode materials, are listed on page 413.

[1] For a discussion of the derivation of this equation see a paper by Saul Dushman, entitled "Electron Emission," *Elec. Engineer.*, July, 1934.

Substance	A	$b_0 = \phi_0/k$
Tungsten (W)	60	52,400
Thoriated tungsten (W + Th)	3	30,500
Barium-strontium oxides ($BaO + SrO$)	10^{-2}	12,000

To illustrate the effect of temperature on the emission let us compute the results in the case of tungsten when operated at, say, 1500°K and at 3000°K. Substituting in Eq. (248) we have, at 1500°K

$$\begin{aligned} I_e &= 60.2 \times (1500)^2 \times 2.718 \times \frac{-52{,}400}{1500} \\ &= 6.79 \times 10^{-7} \text{ amp/cm}^2. \end{aligned}$$

At 3000°K

$$\begin{aligned} I_e &= 60.2 \times (3000)^2 \times 2.718 = \frac{-52{,}400}{3000} \\ &= 13.6 \text{ amp/cm}^2. \end{aligned}$$

Thus we see that by doubling the temperature the emission is increased enormously.

An illustration of the effect of the work function on the emission may be had by computing the emission from a thoriated-tungsten cathode at, say, 1200°K. Again utilizing the Richardson-Dushman relation we have

$$\begin{aligned} I_e &= 3(1200)^2 \times 2.718 \times \frac{-30{,}500}{1200} \\ &= 4.02 \times 10^{-5} \text{ amp/cm}^2. \end{aligned}$$

It will thus be seen that at a lower temperature than the lesser value for pure tungsten in the previous illustration we get a higher output. By the use of such composite emitting surfaces it is, therefore, possible to secure high emission at comparatively low filament temperatures.

It is not our purpose to discuss in detail the character of the various cathodes in current use; it will suffice to add that pure tungsten is used for the filaments of high-power thermionic tubes, while cathodes of the barium-strontium oxide type are extensively used in the small tubes commonly found in radio receivers. Thoriated tungsten is frequently used as a cathode material in medium- and low-powered transmitting tubes. Detailed descriptions of cathode materials and construction will be found in any standard work on electronics or radio communication. An extended discussion of the subject of thermionic emission appeared in

a paper under that title in the October, 1930, issue of *Review of Modern Physics* by S. Dushman.

189. The Effect of the Anode. Later we shall take up in detail the methods by which liberated (free) electrons may be controlled, but at this point it may be well to consider briefly the effect which a second electrode may have on the number of electrons which permanently leave the surface from which they are liberated by thermal means.

In an evacuated space (Fig. 279) the electrons liberated from the cathode C by heat will remain near that electrode unless they are acted upon by an electrostatic or a magnetic field. A cloud of electrons or, as it is called, a **space charge** comes into being. The presence of this body of electrons tends, because of electrostatic action, to inhibit the escape of additional charges. In fact, if the heating continues, a condition will come to obtain where as many electrons will return to the cathode, per unit of time, as escape. Such a situation is analogous to that which exists near the surface of a liquid when a condition of saturation exists.

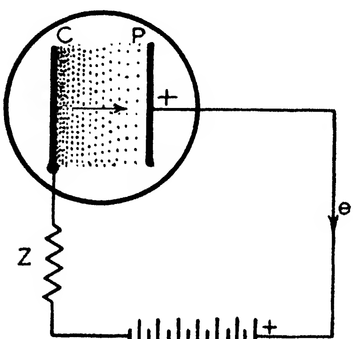


FIG. 279.—Effect of the anode on thermionic emission.

If, however, the second electrode P is made positive with respect to C, as shown, the negative entities constituting the electron cloud will be attracted to this positive member (anode) and thus a **space current** will be established; and this electronic current will continue around the external circuit as indicated. (The conventional current is said to move in the opposite direction.) The question now presents itself as to what effect the potential of the plate (anode) and the distance between the cathode and the anode have on the magnitude of the space current.

In 1911 Child attacked the problem analytically and derived an expression giving the magnitude of the space current in terms of the anode potential and the separation of the two surfaces. His equation,[1]

$$I = \frac{\sqrt{2}}{9\pi}\sqrt{\frac{e}{m}}\frac{V^{3/2}}{d^2}, \tag{249}$$

is based on the assumption of two infinite planes; all electric quantities are in esu. If changed to practical units, the above equation

[1] CHILD, *Phys. Rev.*, **32**, 498 (1911). The development of this equation is also to be found in Professor Chaffee's book "The Theory of Thermionic Vacuum Tubes," p. 68.

becomes

$$I = 2.336 \times 10^{-6} \frac{V^{3/2}}{d^2} \qquad \text{amp/cm}^2. \tag{250}$$

The most significant factor in this relation is the anode (plate) potential term. It is to be noted that, for a given cathode temperature, the space current varies as the $\frac{3}{2}$ power of the potential difference between the cathode and the anode. For this reason the law is often referred to as the $\frac{3}{2}$**-power law.**

This same law was later arrived at independently by Langmuir, who extended the relation to include practical geometrical forms. Langmuir

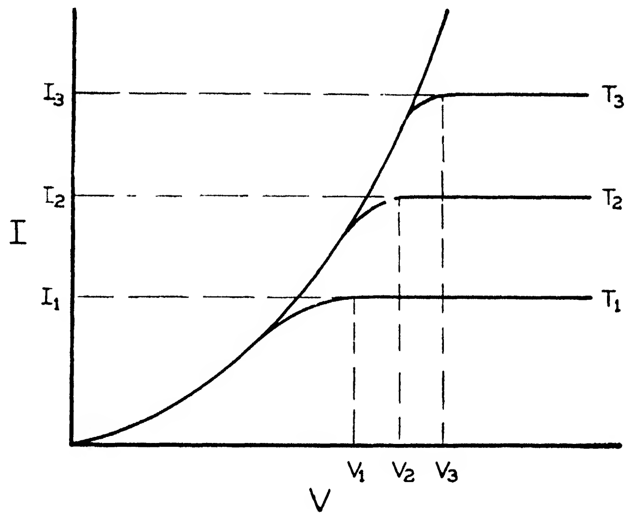


FIG. 280.—Effect of cathode temperature on thermionic emission.

has shown that in the case of two concentric cylinders (cathode and anode both cylindrical) the equation takes the from[1]

$$I = \frac{2\sqrt{2}}{9} \sqrt{\frac{e}{m}} \frac{V^{3/2}}{r\beta^2}, \tag{251}$$

where r is the radius of the anode and β is a constant whose square varies between 0 and 1, increasing in value as the ratio r/a increases, a being the radius of the cathode. If the ratio of r to a is greater than 10, which is often true in practice, β^2 may be taken as unity. In Eq. (249) and (251) e is the electronic charge and m is the mass of the electron. As it stands, Eq. (251) is in esu. When the change to practical units is made,

[1] For the development of this equation see Chaffee, *loc. cit.*, or Appendix *C* in "Physics of Electron Tubes" by Koller.

we have

$$I = 14.65 \times 10^{-6} \frac{V^{3/2}}{r}, \tag{252}$$

where r is in centimeters, V in volts, and I is the space current in amperes **per unit length** of the cylindrical cathode. In practice, both the cathode and the anode may not be cylindrical but Langmuir has shown that the $\frac{3}{2}$-power law holds whatever the shape of the electrodes. The above relation is sometimes referred to as **Langmuir's equation.**

It should be noted in passing that if the plate voltage is high enough, all the electrons that are being liberated at a given cathode temperature will be attracted to the anode, thus giving what is spoken of as the **saturation current.** In Fig. 280 may be seen the plate-potential space-current curves for several different cathode temperatures. At anode potentials above V_1, V_2, and V_3, any increase in plate voltage will not result in an increase in electron current at that particular cathode temperature. I_1, I_2, and I_3 indicate the values of the saturation currents under the conditions indicated.

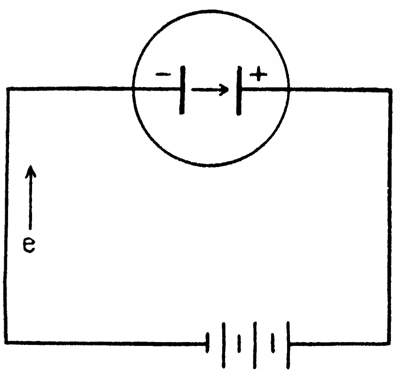


FIG. 281.—Cold-cathode emission.

190. Field Emission. The third method by which electrons may be detached from a substance consists in the establishment of a strong electric field in the region of the surface involved. Neither radiation nor heat have a part in this process. Because thermal energy does not figure in this method of electron liberation, the process is sometimes referred to as **cold-cathode emission,** or **autoelectronic emission.**

Although this method of securing a supply of free electrons is the oldest procedure of its kind known to the art, no generally accepted explanation of the process has as yet been suggested. If two electrodes are sealed in an evacuated enclosure (Fig. 281) and a potential difference of the order of 10,000 or more volts is applied to these terminals an electronic current will pass in the direction indicated in the sketch. This will happen even though the tube is in darkness and is at or below room temperature. The usual explanation given to account for this phenomenon is that the existance of a marked potential gradient in the region of the cathode serves to cause a narrowing of the potential barrier at the emitting surface. A more rational explanation would appear to be that even at relatively low gas pressures a few atoms of gas are still

present. We now know that cosmic rays (Sec. 228) are continually causing ionization effects. It follows then that there will be at least a small number of free electrons, and a corresponding number of positive ions, present in the most attenuated gases. Under the action of the high potential difference these few electrons will acquire high velocities. As a result, ionization by collision occurs. Further, under the circumstances indicated, the resulting positive ions will acquire a relatively high velocity and thereby give rise to electronic liberation as a result of ionic bombardment of the cathode. The early X-ray (Sec. 210) and cathode-ray tubes (Sec. 198) were of the cold-cathode type; and anyone who has used one of the original X-ray tubes knows from experience that a trace of gas was necessary for their successful operation. Today cold-cathode tubes are extensively used in certain types of electronic equipment; and because of improvements in design these units operate at comparatively low anode voltages. In a later chapter we shall consider certain phenomena associated with high-field emission.

191. Secondary Emission. In preceding sections reference has been made to the fact that positive ions may cause electronic emission. The

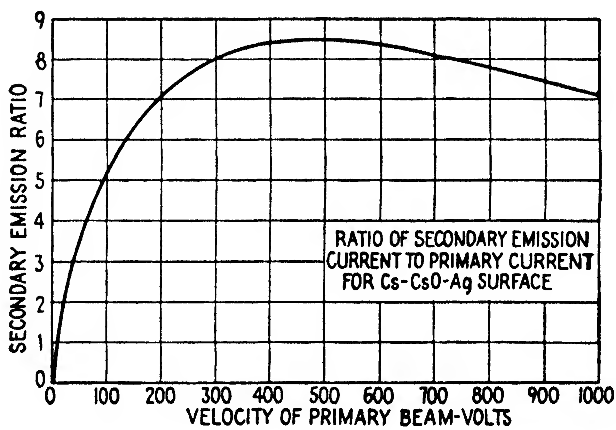


FIG. 282.—Relation between the velocity of primary electrons and the number of secondaries liberated.

term "secondary emission," however, usually refers to the electrons liberated as the result of the bombardment of a substance by primary electrons. If an electron strikes the surface of a substance at a sufficiently high velocity—*i.e.*, if the impinging electron has an energy content of, say, a few hundred electron volts—several electrons may be liberated from the receiving surface. The number of secondary electrons liberated by one impacting electron depends upon (1) the velocity of the primary electron, and (2) the nature and condition of the receiving

surface. The velocity relation, however, is not a linear one. In fact, when the primaries have velocities beyond a certain value, the number of secondaries actually diminishes as the velocities of the primaries are further increased. This is shown by the graph forming Fig. 282, which is reproduced from a paper[1] by Zworykin, Morton, and Walter. The total number of electrons liberated per primary electron may be as high as 10, though the ratio is commonly somewhat less than that. The velocity of the secondary electrons is less than that of the primaries and they leave the bombarded surface in all directions. Though the existence of secondary electrons was known for many years, this type of electronic emission was not put to any practical use until quite recently. In fact, the production of secondary electrons was, and still is, highly undesirable under certain circumstances. In recent years, however, so-called electron multipliers have been designed in which the phenomenon is utilized for important purposes. In such cases special surfaces are prepared which yield a relatively large number of secondaries for each impinging primary. In passing, it should be noted that secondary emission may occur when high-speed electrons impinge on some insulators. In the chapter on electron tubes we shall again refer to the matter of secondary emission.

PROBLEMS

1. Express an electron energy value of 5×10^{-8} erg in mev.

2. Calculate the threshold frequencies for tungsten, magnesium, and caesium.

3. Find the energy required to liberate an electron from a caesium surface. What will be its velocity as it leaves the surface?

4. With what velocity will an electron escape from a sodium surface when illuminated by light whose wave length is 4,400 Å?

5. What area of barium-strontium coated wire would be required to provide an emission of 10 amp when operated at 1200°K?

6. Compare the emission from a pure tungsten filament and a thoriated tungsten filament when both are operated at 1500°K.

7. Compare the space current between a cylindrical cathode and a cylindrical anode when the potential difference between the anode and cathode is 300 volts with what it would be when the potential difference is 500 volts.

[1] Zworykin, Morton, and Walter, "Secondary Emission Multiplier," *Proc. IRE*, March, 1936.

CHAPTER XXV

CONTROL OF ELECTRONS

192. Control of Electrons by Means of a Grid. In Sec. 189 one aspect of the control of free electrons was considered. In that discussion it was pointed out that a positively charged plate, located in the region of the cathode, tended to dissipate the space charge by attracting the electrons constituting that body of charges. The next step in the process of control has to do with the introduction of a **third electrode** into the tube assembly. This third element commonly takes the form of a **grid** which is located relatively near the cathode.

It will assist in arriving at an understanding of the functioning of a grid as a control member if we consider, briefly, the character of the potential distribution within an electronic tube.

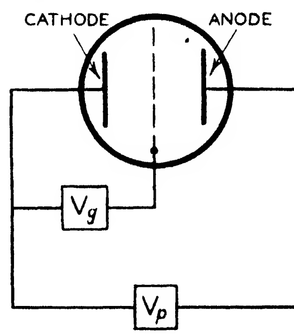


FIG. 283.—Three-element thermionic tube.

Let us assume that we have a three-element unit, which might be represented diagrammatically as in Fig. 283. For simplicity we shall also assume that the cathode and anode are plain parallel surfaces and that the control member (grid) consists of a mesh of fine wires.

First it should be noted that the presence of a space charge changes the potential gradient from what it would be if only a few widely scattered electrons were present in the region between the cathode and the anode. In the latter instance the situation would be as indicated in Fig. 284*a*. In such a case the potential gradient is uniform, as indicated by the straight line OV, and the electrostatic force on each electron is the same at every point between the electrodes, *i.e.*, dV/ds is constant [Eq. (58)].

When a large number of electrons are present, however, the interelectronic forces tend to modify the potential distribution, and the situation becomes something like that pictured in Fig. 284*b*. Those electrons which are nearest the anode are urged **forward** by those immediately behind them; while those near the cathode are retarded by the repellent action of those in front. **The result of these interelectronic forces is to change the slope of the potential curve.** At such a point as p the slope will be relatively great as compared with its value at, say,

p'. Remembering that $\mathcal{E} = dV/ds$, it becomes evident that those electrons near the anode will be acted on by a greater force than those that happen to be in the region of p'.

In the case just cited it was assumed that no grid was present, or, if one were present, that it was maintained at zero potential. **If and when a grid is present,** and maintained at a fixed **positive** potential with respect to the cathode, the slope of the voltage distribution curve will be changed and will appear something like the graph depicted in Fig. 285*a*. The positive potential near the cathode will thus be increased,

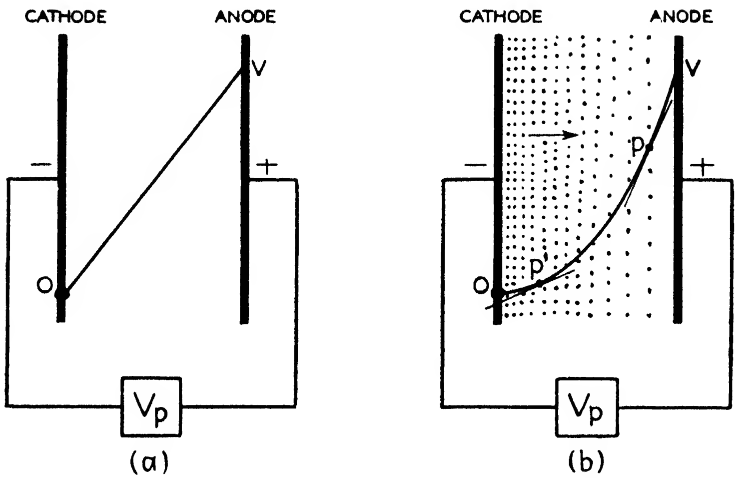


FIG. 284.—Effect of a space charge on the potential gradient in the region between the cathode and the anode.

with the result that the free electrons in that region will be acted upon by a greater force and the space current will accordingly be augmented.

If, on the other hand, the grid is given a **negative** potential, the case will be as diagrammed in Fig. 285*b*. In this event the slope, over a part of its length, will be negative—which means that a repellent force exists in that region—with the result that the accelerating force will be **less** than that due to the plate alone. Indeed, if the grid be made sufficiently negative, its field may completely neutralize the plate's field, thus entirely stopping the movement of the electrons to the anode. In other words, under these circumstances the space current would drop to zero. The negative grid potential which will bring this about is sometimes spoken of as the **cutoff potential.**

It is also important to note that, since the grid is spatially nearer the cathode than is the anode, a unit of potential difference between the grid

and cathode produces a greater effect on the space charge than does a corresponding potential on the plate.

From the foregoing discussion it will be seen that the movement of the free electrons between the cathode and the anode may be brought under more or less complete control by a third electrode, upon which may be impressed various potential magnitudes, either positive or negative in character.

The addition of the third, or control, electrode into an electronic tube was made by Dr. Lee de Forest, on which he was granted a United States

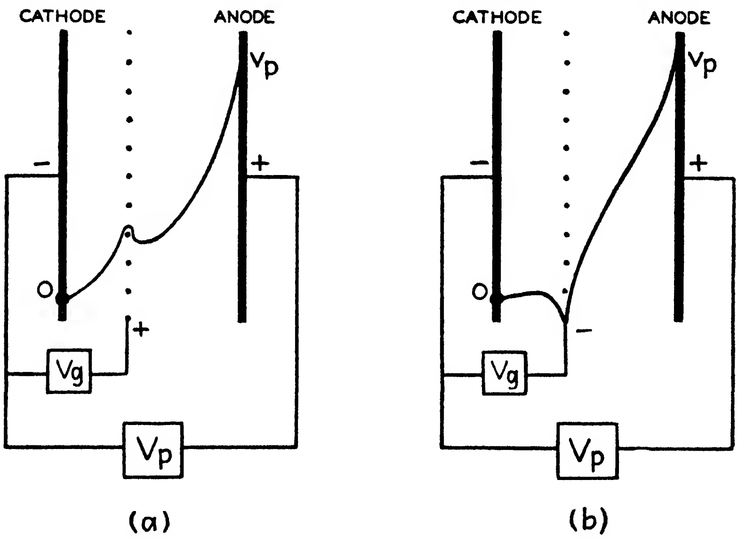


FIG. 285.—Effect of a third electrode on the potential distribution between the cathode and the anode. (*a*) Grid positive with respect to the anode; (*b*) grid negative.

patent in 1908. From time to time various discoveries and inventions have been made which have not only exerted a profound influence on the scientific progress of the age, but have also brought about marked social, economical, and political changes in the world at large. The introduction of the three-electrode tube by Dr. de Forest stands as one of the great contributions to world progress. In later chapters we shall examine various applications of this epoch-making device.

193. Electron Beams. Thus far in dealing with the control of electron motion we have been considering electron groups having sizable dimensions. We are now to examine those cases which involve a thin pencil, or beam, of charges, the units of which are given high linear velocities by the application of suitable accelerating potentials, and which may also be acted upon by electrostatic or magnetic fields.

A narrow stream of electrons may be produced by an organization such as that sketched in Fig. 286. The housing is of glass, and the region within is evacuated. An electrically heated filament C yields a supply of thermions which are attracted toward the anode A as a result of the accelerating potential supplied by the d-c source V. In its simplest form the anode (accelerating electrode) consists of a hollow metallic cylinder in both ends of which is a round hole whose diameter is of the order of a millimeter. This cylindrical structure serves as a collimating device. All electrons issuing from the cathode will be stopped except those constituting a small bundle, the units of which are moving **in parallel paths.** As a result of this collimating action there becomes available, at the right of the anode, a so-called **cathode ray.** The electrons constituting this slender beam, as it issues from the collimator, will be moving with a velocity which will be determined by the accelerating potential, *i.e.*, by

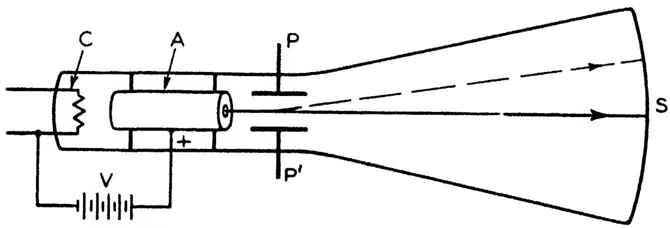


FIG. 286.—Essentials of an electron-beam tube.

the potential difference that exists between the cathode and anode. Unless acted upon by some electrostatic or magnetic force this velocity will remain constant until the electrons impinge upon the end of the tube. If the inside surface of the broader end S of the tube be coated with one of several materials (say, willemite) fluorescence will occur at the point where the electrons strike the sensitive layer. That part of the organization described above which has to do with the production and collimation of the stream of free electrons is often referred to as an **electron gun,** and the assembly as a whole as **a cathode-ray tube.**

In passing, it should be recorded that it is possible to project cathode rays outside the tube in which they originate. By arranging a thin aluminum window in the end wall of a cathode-ray tube, Lenard, in 1894, found that the electrons constituting the cathode stream would pass through the metal foil and continue into the air beyond. These escaping electrons appear to retain their original properties, though they soon lose their energy as they encounter the atoms of the atmospheric gases. A stream of electrons existing outside the tube in which it originates is spoken of as a **Lenard ray.** In Sec. 231 we shall have occasion to consider

an interesting and significant phenomenon which appears when electrons pass through a thin sheet of metal, such as that used in a Lenard tube.

194. Electrostatic Control of an Electron Beam. In the early part of our discussions it was pointed out that a stream of electrons constitutes an electric current. It was also noted that moving electrons give rise to a concomitant **magnetic field.** We have also seen that electrons at rest or in motion are surrounded by an **electrostatic field.** The foregoing statements hold whether the electrons are moving along a conductor or **in free space.** It therefore becomes evident that an electron beam, as produced by an electron gun in a cathode-ray tube, may be brought under the influence of an electrostatic or a magnetic field in such a way as to modify and to control the trajectory of the electrons constituting the beam.

Let us consider first the application of an electrostatic field. If a difference of potential is established between the plates P and P' (Fig. 286), making P, say, positive with respect to P', the moving electrons will be acted on by an electric force whose direction is at right angles to their line of motion. Owing to the attraction of the positive plate and the repulsion due to the negative member, deflection will occur while the electrons are under the influence of this field, and after emerging from the region of the field the electrons will continue to move in straight and parallel paths, as indicated by the dotted line. Thus, by applying a potential difference to the deflecting plates, the electron beam may be made to function as a long massless pointer. If an alternating potential difference is applied to the plates, the end of the beam will sweep up and down and thus give rise to a straight visible trace on the fluorescent screen. Obviously, one use of such a device would be as a voltmeter. In connection with this and other uses to which the cathode-ray tube is put it becomes necessary to know the relation between the accelerating voltage and the deflecting potential. In order to arrive at this relation it is necessary to be able to compute the linear speed of the electrons as they arrive in the region between the plates, and also the speed given to the electrons by the transverse electric field.

The speed of the electrons as they reach the deflecting plates may be arrived at by equating their kinetic energy to the work done on the electrons by the cathode-anode field, thus

$$\tfrac{1}{2}mv^2 = eV,$$

where V is the potential difference between the cathode and the anode. From the above we see that the velocity is given by

$$v = \sqrt{\frac{2eV}{m}}. \tag{253}$$

This expression gives the longitudinal **speed** not only at any point within the plates but also at any and all points between the plates and the screen. While the **linear speed does not change** as an electron passes through the deflecting field, **its velocity does change**; the direction factor of the vector changes, but not the magnitude factor. Since we know that the value of the ratio e/m is 1.765×10^7 emu/gm, and that 1 volt equals 10^8 abvolts, the last equation reduces to the form

$$v = 5.95 \times 10^7 \sqrt{V} \qquad \text{cm/sec.} \tag{254}$$

If the cathode-anode potential difference were known to be, say, 1,000 volts, the speed of the electrons would turn out to be 1.88×10^9 cm/sec.

An expression giving the **tranverse** speed can be deduced by making use of Newton's second law, which in algebraic form is

$$F = ma \tag{i}$$

Equation (4) is, for our present case,

$$F = \mathcal{E}e; \tag{ii}$$

and, by the use of the relation expressed by Eq. (59), (ii) becomes

$$F = \frac{V'}{s} e, \tag{iii}$$

where V' is the potential difference between the plates and s the distance

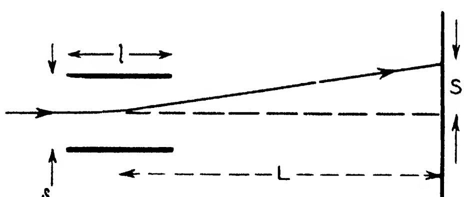


FIG. 287.—Control of an electron beam by means of an electrostatic field.

between them as indicated in Fig. 287. Combining (i) and (iii) we get

$$a = \frac{V'e}{ms}. \tag{iv}$$

The transverse speed will be the product of this acceleration and the time during which it acts. In this case

$$t = \frac{l}{v'}, \tag{v}$$

where l is the length of the charged plates and v' the transverse speed.

Combining (iv) and (v) we have

$$v' = \frac{V'el}{msv}. \tag{255}$$

Each electron, then, has impressed upon it two velocities, viz., v and v', at right angles to one another. In a given interval of time, the corresponding displacements will be proportional to these velocities. Hence we may write,

$$\frac{S}{L} = \frac{v'}{v} = \frac{V'el}{msv^2},$$

which, by the use of Eq. (253), reduces to

$$S = \frac{1}{2}\frac{lLV'}{sV}. \tag{256}$$

An examination of the above relation discloses the fact that the linear displacement of the spot on the fluorescent screen will depend upon the length and spacing of the plates and upon the accelerating potential. By changing Eq. (256) to the form

$$\frac{S}{V'} = \frac{1}{2}\frac{lL}{sV}, \tag{257}$$

we have an expression for what may be called the electrostatic or voltage sensitivity of the tube. This term is an index of the screen deflection one may expect per volt potential difference on the deflecting plates. The reciprocal of the sensitivity is sometimes used, and is referred to as the **deflection factor.** In practice, a tube sensitivity of the order of 1mm/volt is common. This sensitivity can be greatly increased by the use of amplifiers, as will be explained later.

195. Magnetic Control of an Electron Beam. The control of an electronic beam by means of a magnetic field finds a number of important applications. It will, accordingly, be useful to have available a relation by the use of which one may compute what the deflection of the beam will be under a given set of circumstances.

It has already been pointed out that a moving charge is equivalent to an element of current. It therefore follows that Eq. (152) is applicable in the case before us. That relation gives the mechanical force experience by an element of current which is located in a magnetic field whose direction is normal to the direction of the current element. The equation is

$$F = HI \qquad \text{dyne-cm.} \tag{i}$$

Since, in this case, $I = e/t$ we may change (i) to the form

$$F = H\frac{e}{t}. \qquad \text{(ii)}$$

In Fig. 288, it will be seen that the length of the current element involved is l; hence

$$F = He\frac{l}{t} = Hev \quad \text{dynes,} \qquad \text{(iii)}$$

where v is the velocity with which each electron enters the magnetic field. Since the force given by (iii) is normal to the initial direction of motion, the magnitude factor (the speed) will not change. However, the radial acceleration, as in the corresponding case in simple mechanics, will be given by

$$a = \frac{v^2}{r}, \qquad \text{(iv)}$$

where r is the radius of the curved path followed by the electron while in the magnetic field. The centrally directed force acting on each electron will therefore be

$$F_c = \frac{mv^2}{r} \quad \text{dynes,} \qquad \text{(v)}$$

where m is the mass of the electron. From the fundamental laws of mechanics it therefore follows that we may equate the two forces designated by F and F_c, as given by (iii) and (v). This gives

$$Hev = \frac{mv^2}{r}. \qquad \text{(vi)}$$

Solving (vi) for r we get

$$r = \frac{mv}{He} \qquad \text{(vii)}$$

as the radius of the circular path described by the electrons **while they are passing through the magnetic field.** If now we substitute for the initial velocity its equivalent, as given by Eq. (253), we have, after reducing,

$$r = \frac{1}{H}\sqrt{\frac{2mV}{e}}. \qquad (258)$$

This relation, then, gives the radius of the trajectory **within a uniform transverse magnetic field.** Equation (258) should be kept in mind when we come to consider the fundamentals of the mass spectrograph (Sec. **207**).

After leaving the field the electrons will follow a straight path to the fluorescent screen. The magnitude of the resulting deflection S can be deduced from the factors involved, as follows. Referring to Fig. 288, if the angle θ is small, as is usually the case in practice,

$$\theta \doteq \frac{S}{L} = \frac{l}{r}.$$

Substituting for r its equivalent, as given by Eq. (258), we have

$$\frac{S}{L} \doteq lH\sqrt{\frac{e}{2mV}}.$$

Solving for S,

$$S = lHL\sqrt{\frac{e}{2mV}}. \tag{259}$$

Thus we have an expression for the deflection produced by a given magnetic field when acting on a beam of electrons which are moving under the

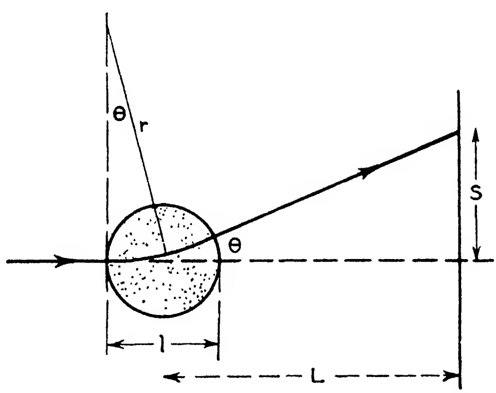


FIG. 288.—Control of an electron beam by means of a magnetic field. The dotted region represents a cross section of the magnetic field.

influence of an accelerating potential of V abvolts. In order to produce a uniform field recourse would be had to a set of Helmholtz coils, one coil being positioned on each side of the tube.

196. Determination of the Ratio e/m. The cathode-ray tube affords a convenient means by which it is possible to determine the numerical value of the ratio e/m. An expression giving the ratio e/m may be derived from Eq. (258). It will be

$$\frac{e}{m} = \frac{2V}{H^2r^2}. \tag{260}$$

Here, then, we have a relation which gives, in terms of readily measurable quantities, what is perhaps the most important ratio in the realm of physics. As the equation stands, V, the accelerating potential, is in abvolts. Changing to volts, this expression becomes

$$\frac{e}{m} = \frac{2V \times 10^8}{H^2r^2}. \tag{261}$$

The value of the accelerating potential can be readily and accurately determined. The field H can be computed from the winding of the Helmholtz coils and the current they carry, or it may be measured directly by means of a fluxmeter. The radius of the curved part of the path is not so easy to determine accurately. In Sec. 195 it was pointed out that when θ (Fig. 288) is small, $\theta \doteq l/r$. It therefore follows that

$$\frac{l}{r} = \frac{S}{\sqrt{S^2 + L^2}},$$

which gives

$$r = l\sqrt{1 + \frac{L^2}{S^2}}. \tag{262}$$

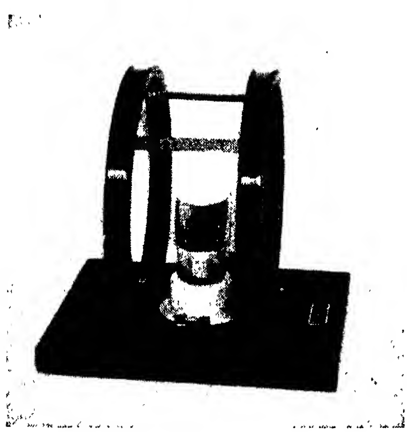


FIG. 289.—Apparatus by means of which the ratio e/m may be determined.

There is now available a specially constructed cathode-ray tube and coil assembly by means of which the factor r may be easily determined. An illustration of the setup is shown in Fig. 289. The tube is so designed that the cathode beam may be bent into the form of a semicircle of known radius. This equipment yields only approximate results, but it illustrates the fundamental principles involved in such a determination.

197. Properties of the Cathode Rays. It is of interest to note that the cathode-ray tube was developed before the nature of the rays was understood. The electrical discharge through gases at reduced pressure was extensively studied by Sir William Crookes, who reported on his observations in a paper in 1879. Several of the properties of what we now designate as cathode rays were noted by Crookes. He felt that he was dealing with what might be spoken of as an ultragaseous state of matter; Crookes referred to it as a "fourth state" of matter. For several years a controversy took place over the subject of the nature of this discharge. It was agreed that something was emitted from the cathode. The German school of thought held, more or less, to the view that the

rays consisted of some sort of wave disturbance; while the English physicists felt that they were dealing with some type of electrified particles.

Jean Perrin, a professor at the University of Paris, appears to have been the first to investigate this question experimentally. As a result of his studies he arrived at the conclusion (1895) that the cathode rays were made up of **carriers of negative electricity,** rather than consisting of some form of radiant energy. In 1897 Sir J. J. Thomson repeated and extended Perrin's experiments. Using a tube whose design (Fig. 290) was the forerunner of the modern cathode-ray tube, Thomson confirmed the findings of Perrin, and also established the fact that **the electrical entities constituting the beam are the same regardless of the nature of the cathode material or the character of any residual gas in the tube.**[1]

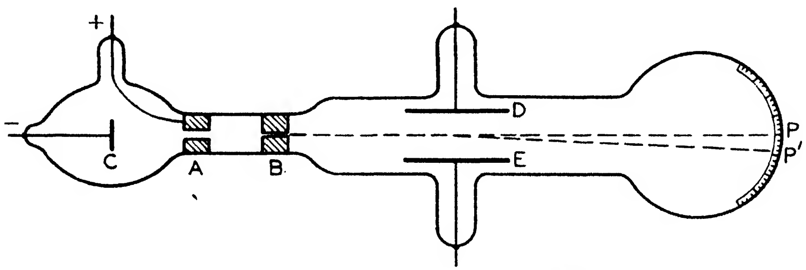


FIG. 290.—Diagram of the tube used by Thomson in his original research on the nature and properties of electrons.

Thomson called these negatively charged entities corpuscles. He established the fact that they possess **mass** as well as a charge, and by applying an electric and a magnetic field simultaneously (one neutralizing the effect of the other) he arrived at a value for the ratio of e/m. Thomson's value was 7.7×10^6 emu/gm, the velocities with which he worked being of the order of 2.2×10^9 to 3.6×10^9 cm/sec. As the result of later investigations by Lenard, Kirchner, Dumington, and others, we now know that the e/m ratio has a value of $1.760 \pm 0.0025 \times 10^7$ emu/gm.

As a result of the work of many investigators, we know that the corpuscles of Thomson are **not** charged material particles, but that they are **elemental negative charges,** which we have come to call **electrons.** As we have already seen, the magnitude of this elemental charge is $(4.803 \pm 0.005) \times 10^{-10}$ esu or 1.591×10^{20} emu. From the known values of e/m and e the mass of the electron can be computed; it is, at moderate velocities, 9.02×10^{-28} gm.

The properties of the cathode rays may be briefly summarized as follows:

[1] THOMSON, J. J., *Phil. Mag.*, **44,** p. 293, (1897).

1. The electrons forming the stream leave the cathode at right angles to its surface.
2. They travel in straight lines.
3. They possess kinetic energy, the magnitude of which depends upon their velocity, which in turn depends upon the value of the accelerating potential.
4. They are deflected by an electrostatic and a magnetic field.
5. Owing to their kinetic energy they will ionize a gas.
6. They will produce fluorescence when they impinge on certain substances.

In concluding our present discussion of cathode rays, it is well to note the fact that the ratio e/m is not strictly constant; it has been found that the value of this ratio is, to some extent, a function of the velocity at which the electron is moving. There is no reason to believe that the magnitude of the electronic charge ever changes. It therefore follows that it must be the mass which is the variable. Such a conclusion at once raises an important question. Is it correct to assume that an electron actually exhibits the universal property which we commonly designate by the term mass? It would be beyond the scope of this book to discuss in detail the question of electronic mass and its dependence upon velocity. However, it may be observed in passing that it is an established fact that the mass of an electron actually **increases** as its velocity approaches the speed of light.

We have already seen that when an electron is caused to move, a concomitant magnetic field comes into being. In order to establish this attendant field, energy must be expended. Further, in order to change the velocity of the electron—*i.e.*, to produce acceleration—a definite force is required. A consideration of these and related aspects of the case has led some physicists to feel that the mass of the electron is wholly magnetic in character and, in fact, that all mass is essentially electromagnetic in nature.

Newton originally gave the law relating force and mass in the form

$$F = \frac{d(mv)}{dt} = v\frac{dm}{dt} + ma.$$

If mass were independent of velocity the above expression would reduce to the more familiar form:

$$F = ma.$$

If, however, one begins with the first of the above forms, it may be shown that

$$m = \frac{m_o}{(1 - (v^2/c^2)^{1/2}},$$

where m is the mass when the particle has the velocity v, m_o its mass when at rest or moving at relatively small speeds, and c the velocity of light.[1] On this basis the mass of an electron may be expected to increase as the velocity increases; and there is reliable experimental evidence in support of the validity of such a relation. Indeed, on the basis of the theory of relativity, it turns out that **all mass,** regardless of its nature, will vary in conformity with the equation above given; and such has been found to be the case. As we shall see in a later section, it is now possible to give electrons a velocity which is within a fraction of 1 per cent of the velocity of light. In such a case the change in mass is appreciable and must be considered when dealing with high speed particles.

For a more complete discussion of the relation of mass to velocity the reader is referred to "The Particles of Modern Physics" by J. D. Stranathan, pp. 129 *ff*.

198. The Modern Cathode-ray Tube. In 1897, Ferdinand Braun designed a tube that made use of the cathode stream as a means of study-

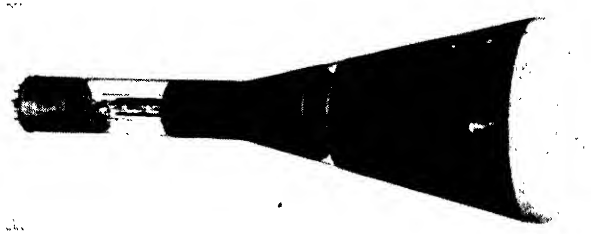


FIG. 291.—Modern cathode-ray tube. Dark portion is an internal electrostatic screen. (*Courtesy of Allen B. DuMont Laboratories, Inc.*)

ing varying potentials and changing magnetic fields. The tubes used by Thomson in England and Kaufmann in Germany were more or less identical with the original Braun unit. The early Braun tubes, as they have come to be called, consisted of a flat cathode from which electrons were liberated by the high-field method, and collimation was effected by causing the beam to pass through a cylindrical channel of very small diameter. The field-emission method of securing free electrons necessitated the application of high potential differences—potentials of the order of 50,000 volts, which were supplied by static machines or induction coils. The hot-cathode tube was introduced by Wehnelt in 1905, but it

[1] This relation is often written in the form

$$m = \frac{m_o}{(1 - \beta^2)^{1/2}},$$

where $\beta = v/c$. The origin of this equation is commonly ascribed to Lorentz.

was not until the early twenties that a practical low-voltage unit was developed. In recent years various improvements have been made until today we have available a modern version of the Braun tube that has come to be one of the most important scientific tools known in research work and in practical engineering. The cathode-ray tube is to the engineer what the stethoscope is to the physician. A modern tube is shown in Fig. 291, and a diagrammatic sketch of a representative tube structure and connections appears as Fig. 292.

An electrically heated wire (Fig. 292) serves to heat a small surrounding metal cylinder, the flat end of which is coated with a material which emits electrons copiously when heated to a moderate temperature. The coated surface functions as a cathode. The cathode structure is in turn surrounded by a metal cylinder, one end of which is closed except for a very small aperture directly in front of the cathode-emitting surface.

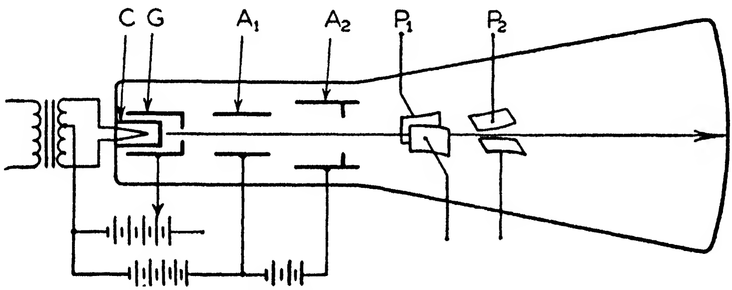


FIG. 292.—Diagrammatic sketch of the essential components of a cathode-ray tube.

The control member G serves the purpose of an electron gate, and is analogous to the diaphragm in a camera or a microscope. A relatively small negative potential impressed on this component will control the number of electrons that pass through the aperture and thus go to form the electron stream. Since the intensity of the fluorescent spot on the screen depends upon the number of electrons which impinge on the surface per unit time, the intensity can be controlled by modifying the negative potential applied to this so-called grid member.

Again referring to the diagram (Fig. 292), it will be noted that there are two anodes, A_1 and A_2. These two anodes serve two purposes: (1) they both act to accelerate the electrons; and (2) taken together they act as an electron lens, *i.e.*, they serve to bring the electron beam to a focal point on the screen. Anode A_1 is maintained at a potential of several hundred volts above that of the cathode, thus giving velocity to the electrons and, because of its mechanical structure, acting to collimate the beam.

As the electrons emerge from the last aperture in the first anode A_1

they will, if left to themselves, **tend to diverge** because of their mutual repulsion. In order to compensate for this, and to concentrate all electrons in any cross section of the beam on the smallest possible area of the screen, recourse is had to electrostatic focusing. This is brought about by means of the two anodes acting conjointly. The second anode A_2 is a metal cylinder whose diameter is somewhat larger than that of A_1, and is maintained at a potential of the order of 1,000 to 5,000 volts higher than the cathode. Under these conditions the electrostatic field will have the form pictured in Fig. 293, with the result that any radial acceleration is neutralized and the beam is brought to a sharp focus on the screen S. The size of the focal point can be controlled by changing the ratio of the positive potential of A_1 and A_2. Thus we have what, in effect, is the equivalent of a lens whose focal length can be adjusted at

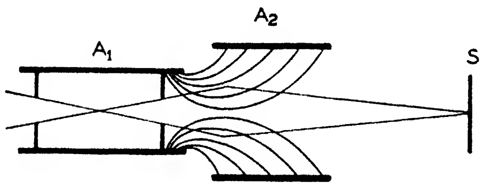


FIG. 293.—Focusing of an electron beam by means of an electrostatic field.

will. In the field of electron optics such a combination is referred to as a **double-cylinder lens.**

The fluorescent screen consists of certain chemical substances deposited on the inside of the large end of the tube. Among the chemical bodies used are: zinc silicate (Zn_2SiO_4), giving a green trace; calcium sulphide (CaS), giving a color which is nearly white; and calcium tungstate ($CaWO_4$), yielding a blue trace. The fluorescent traces which appear on the screen may be observed visually or photographed, using a fast lens and a fast, contrast plate or film. If the instrument is to be used principally for the purpose of making photographic observations, it is called an **oscillograph.** If used for visual observations it is referred to as an **oscilloscope.** Actually, there is no fundamental difference in the two devices, the character of the phosphor used in making the screen being the only variation.

The cathode-ray oscillograph is used chiefly in connection with the study of potential and current wave forms and transient electrical phenomena. This involves provision for a time axis. The time axis is usually provided by applying a "sweep" potential to the horizontal plates. By means to be discussed later, it is possible to produce a potential whose wave form is of the character shown in Fig. 294. It will be noted that the potential rises from zero to some predetermined value,

after which it instantaneously drops to zero. When this potential is applied to the horizontal plates (P_1, Fig. 292) the electron beam will be caused to sweep horizontally across the screen and return to its original position in an exceedingly short interval of time, and continue to repeat this operation. Owing to the persistence of vision, and a certain amount of lag in the screen fluorescence, a horizontal fluorescent line is thus caused to appear on the screen. If now, a variable potential whose wave form it is desired to study is applied to the vertical plates P_2 a trace of the wave will be outlined on the screen. Instead of using a "saw-toothed" voltage on the horizontal plates two sinusoidal voltages may be impressed on the two sets of plates and their combined effect noted. Thus, two frequencies may be compared very accurately, and phase

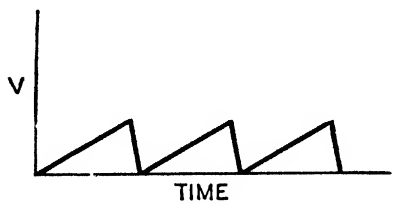


Fig. 294.—Wave form employed in the sweep circuit of a cathode-ray assembly.

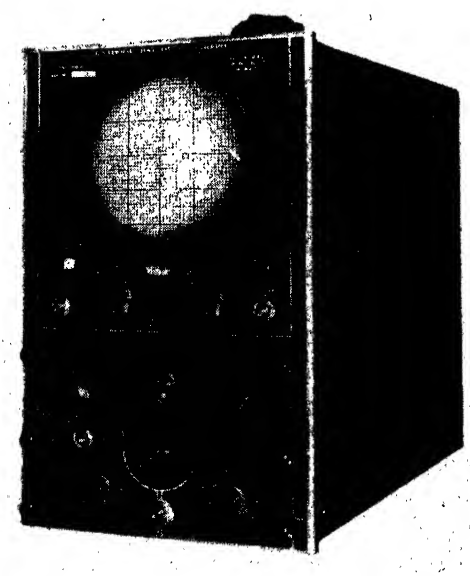


Fig. 295.—Modern cathode-ray oscillograph. (*Courtesy of Allen B. DuMont Laboratories, Inc.*)

differences studied. In the field of acoustics the oscillograph has also come to be extensively utilized. Sound waves can be converted into electrical currents, and these in turn can be investigated by means of such a cathode-ray assembly. From what has been said it will be evident

that the cathode-ray oscillograph is an exceedingly versatile piece of research equipment. It is because this device is so widely used in both research and industry that we have treated the subject somewhat at length. In Fig. 295 is to be seen an illustration of a modern oscillograph. The knobs on the front panel are for the purpose of making the potential adjustments mentioned above, and for other adjustments which add to the convenience of operation. In Fig. 296 is reproduced the oscillographic records of a simple and a complicated voltage-wave form. If the reader is interested in high-speed oscillographic work he should consult a paper by Messrs. Kuehne and Ramo which appeared in the June, 1937, issue of *Electrical Engineering*. The beginner in the use of the oscillograph may find it useful to consult a book by J. F. Rider entitled "The Cathode-ray Tube at Work."

199. The Electron Microscope. Because of the nature of light it is physically impossible to attain a magnification greater than something

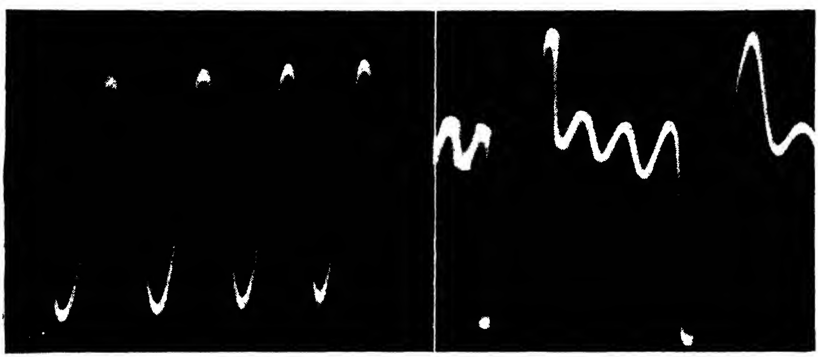


FIG. 296.—Records made by means of a cathode-ray oscillograph.

like 2,500 diameters by means of a microscope when using light waves. By making use of free electrons in a specially designed cathode-ray tube it is now possible to secure magnifications greater than 10 times the value mentioned above. The method by which this end is attained involves the principles of what is now referred to as electron optics.

The theoretical basis of modern electron optics dates back to the early part of the nineteenth century. It was Hamilton who advanced the idea that a point mass when moving through a potential field behaves in the same manner as does a light ray in passing through optically refracting mediums. At that time, of course, we knew nothing about electrons, or electron mechanics. However, as a result of the pioneering work of Perrin, Thomson, Kaufmann, Braun, and others, it became evident that in the electron we have what amounts to a point mass; and that by applying Fermat's principle of least time, as applied to a train of light

waves, and Lagrange's principle of least action, as it applies to a mechanical movement, it would be possible to treat the motion of an electron through an electrostatic or a magnetic field as though it were a train of light waves. Such an analytical procedure involves the assumption that

$$n = k\sqrt{V},$$

where n is the refractive index, V the potential function of the field, and k a constant. According to the Newtonian method of explaining the refraction of light which takes place at an optical interface, we have the relation

$$\frac{n}{n'} = \frac{\sin i}{\sin r},$$

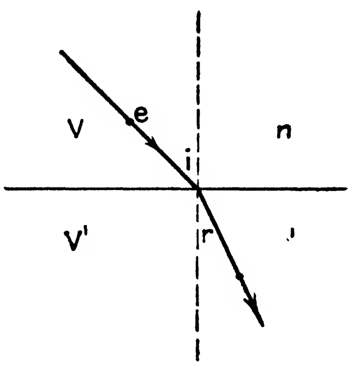


FIG. 297.—Refraction of an electron beam.

where n and n' represent the refractive indices of the two mediums involved, and i and r the angles of incidence and refraction, respectively. Applied to the electronic case, the situation would be as sketched in Fig. 297. Let us assume that an electron e is moving through an electrostatic field where the potential is V, and that it passes into a region where the potential is V', V' being greater than V. Let the electronic refractive indices be n and n', as indicated.

We have shown [Eq. (253)] that the velocity of an electron when in an electric field is proportional to the square root of the potential. Hence we may write

$$\frac{v}{v'} = \frac{\sqrt{V}}{\sqrt{V'}},$$

which shows that the velocity in the second region will be greater than in the first. The incident velocity v may be thought of as having two components, one of which is normal to the interface and the other parallel to it; and the same will be true of v'. But the tangential components are not changed by the change in the field. It accordingly follows that

$$\frac{\sin i}{\sin r} = \frac{v'}{v} = \frac{n'}{n} = \frac{\sqrt{V'}}{\sqrt{V}}.$$

The part of the foregoing relation that concerns us is

$$\frac{\sin i}{\sin r} = \frac{\sqrt{V'}}{\sqrt{V}}, \tag{263}$$

which shows that the path of each of the electrons constituting a cathode beam will be deflected toward the normal when entering a region where the potential is relatively high, and vice versa. There is one marked difference between the optical and the electronic case. Except in the case of the retina of the eye and the atmosphere, the refractive index is more or less constant throughout a given medium; in other words, sharply defined boundary conditions obtain. In an electrostatic field the potential gradient may not be constant. Hence the change in the direc-

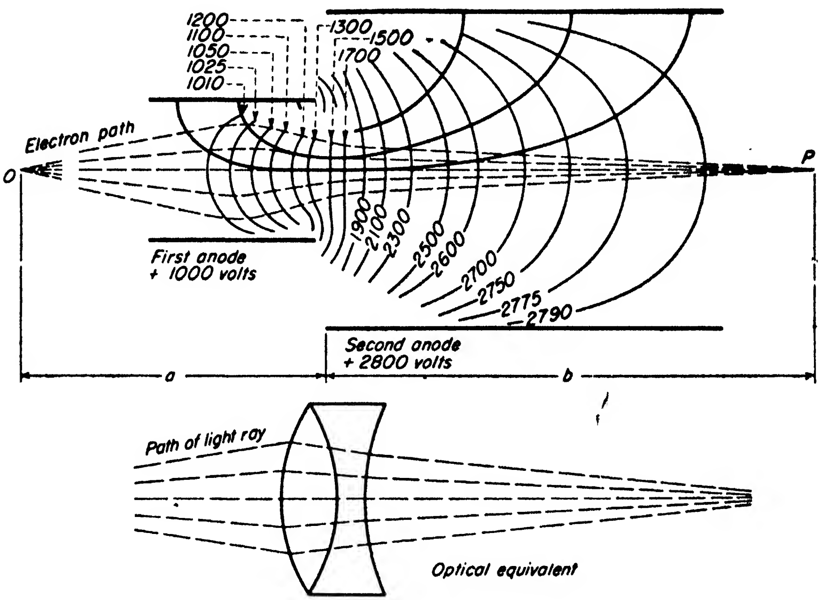


FIG. 298.—Diagrammatic representation of the electric field and the electron paths within a two-cylinder lens. The lower sketch shows the optical equivalent. (*From a paper by V. K. Zworykin, courtesy of J. Franklin Inst., Vol.* 215, *May*, 1933.)

tion of the electronic movement will, in general, not take place abruptly, but gradually. If we redraw the diagram shown in Fig. 293 so as to include the **equipotential lines,** the situation would be as sketched in Fig. 298. Having the equipotential lines, we can apply Eq. (263) by a step-by-step process and thus determine the trajectory of the electrons making up the beam. The equipotential lines correspond to the bounding surfaces in a composite lens. The optical equivalent of the double-cylinder electron lens is to be seen in the same illustration. After coming to a "focus" at P the electrons, unless stopped, will again diverge, as in the case of light waves. By the use of a group of such electrostatic

electron lenses it is possible to assemble an "optical" instrument, such as a microscope or a telescope.

It is also possible to utilize a **magnetic field** as an electron "lens." This fact was first shown by Wiechert in 1899. When so used the axis of the electromagnet coincides with the axis of the cathode beam. A form of magnetic lens currently used is sketched in Fig. 299. This consists of a winding inside an iron shell which has an annular opening. Such an arrangement gives rise to two annular poles as shown. The lines of magnetic force and equipotential lines (magnetic potential) will have the general form sketched. In our discussion of electrostatic lenses we saw that equipotential surfaces correspond to optical surfaces. In the case before us it will be seen that those electrons which travel along the path b will be everywhere normal to the equipotential surface, hence no deviation will occur. However, those that follow the trajectories indicated by a and c will be deviated as shown. This deviation is a somewhat more involved process than indicated by the simple diagram. Actually, those electrons which are not following a path which is strictly parallel to the magnetic field are caused to describe helical paths, rejoining the axial electrons at a point (focal point) which is determined by the value of the accelerating voltage and strength of the magnetic field.[1] The final result is that the electron beam will converge to a point P, and if not stopped, or again acted upon, will again diverge, as in the corresponding optical case. By utilizing a group of such magnetic lenses it is possible to assemble what amounts to a microscope—a device in which electrons take the place of light waves, and coaxial magnetic fields function as optical lenses.

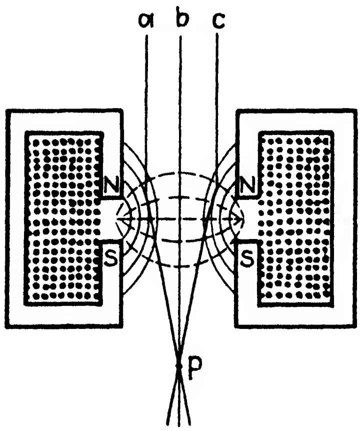


FIG. 299.—Focusing of an electron beam by means of a magnetic field.

[1] For a clear account of the theory of magnetic focusing the reader is referred to a paper by V. K. Zworykin entitled "On Electron Optics," which appears in the May, 1933, issue of the *Journal of the Franklin Institute*, p. 535. The October, 1944, number of the *Journal of Applied Physics* contains several papers on the history and applications of the electron microscope. See also an article in the June, 1945, issue of the *Journal of the Franklin Institute*, by R. G. Picard on "New Developments in Electron Microscopy."

Practical electron microscopes are now available, some of which utilize electrostatic lenses and others magnetic lenses, the latter type being the more common. A diagrammatic sketch of an assembly of the magnetic type is shown in Fig. 300. The drawing is more or less self-explanatory. A hot, self-emitting cathode is used. The electrons are given a high linear velocity by means of an anode potential of the order of 30,000 to 100,000 volts. The object to be examined is supported on a suitable stage at the point shown. The specimen sections must be exceedingly thin; some electrons are stopped by the object under examination while others pass through, as in the case of light. The transmitted electrons are brought to a conjugate focal point by the magnetic lens, which functions as an objective, thus forming a first or primary image in the plane indicated in the diagram. This image is a shadow of the object; it may, if desired, be further magnified by the third magnetic lens, thus forming a secondary or final image on a fluorescent screen or on a photographic film located in the base of the apparatus. Provision is made for viewing the final image through one or more windows in the side wall of the base. The entire interior of the housing is evacuated. In the latest models provision is made for introducing the specimen and the photographic film without disturbing the complete vacuum.

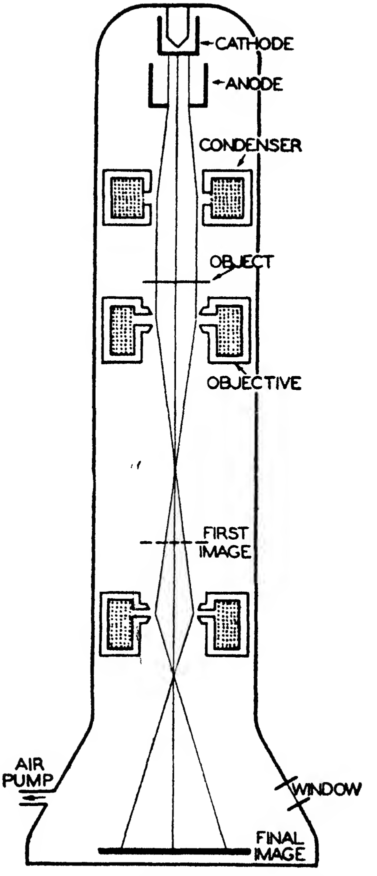


FIG. 300.—Diagrammatic sketch of the essentials of an electron microscope.

Figure 301 shows a recent model of an electron microscope, and Fig. 302 is a reproduction of a representative electronic skiagraph. The photograph shows the structure of diamond dust. A detailed study of the original picture indicates that there was some penetration by the electrons, and that Bragg reflections occurred. In this particular case the original electron optical magnification was 5,600. The reader will find an informative paper in the December, 1944, issue of the

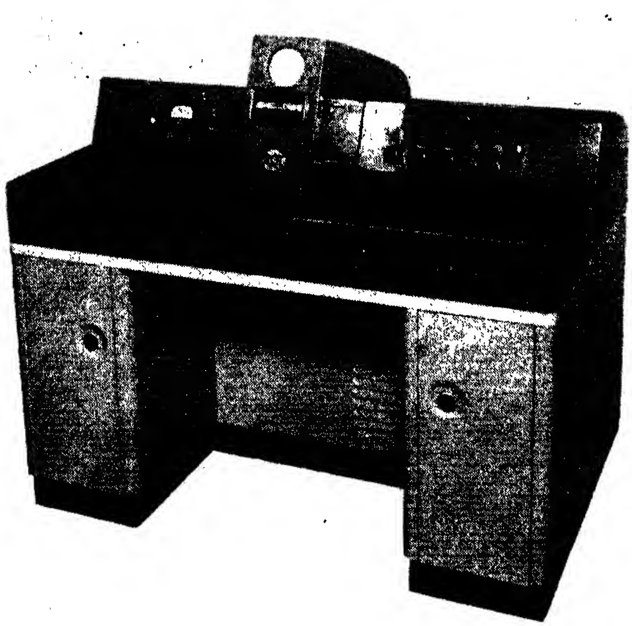


FIG. 301.—Console model of an electron microscope. (*Courtesy of Radio Corporation of America.*)

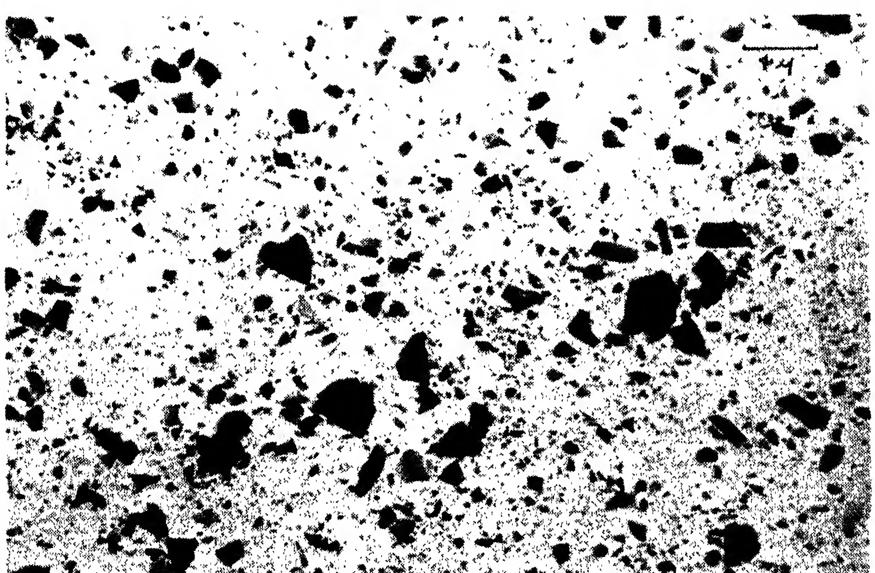


FIG. 302.—Electronic skiagraph. (*Courtesy of E. F. Fullam, Research Laboratory, General Electric Company. Sample from the Schenectady Works Laboratory.*)

General Electric Review, by J. B. Bensen, in which some of the more recent models of electron microscopes are described. The final paragraph is particularly worthy of note.

By means of an electron microscope it is possible to secure, within the instrument itself, a magnification of the order of 25,000 diameters, with resolution which is considerably higher than that attained by means of an ordinary microscope. The resulting definition permits the optical enlargement of such electron skiagraphs, thus making it possible to secure over-all magnifications of the order of 100,000. It seems probable that it will soon be possible to examine molecular structure visually. The development of the electron microscope makes available a research tool of tremendous value in the fields of biology and the industrial arts.

PROBLEMS

1. Electrons are caused to enter an electrostatic field between two charged plates that are 1 cm in length and spaced 0.5 cm apart. A fluorescent screen is located 16 cm from the center of the deflecting plates. The accelerating voltage is 1,000. What will be the voltage sensitivity of the assembly?

2. An electron whose energy is 2,000 ev enters a magnetic field of 100 gausses, the field being normal to the electron path. What will be the radius of the trajectory as the electron moves through the field? What would the radius be if the electrical entity was a proton, all other conditions being the same?

3. A mixture of α-particles and deuterons are subjected to an accelerating potential of 1,000 volts, and caused to pass through a magnetic field whose strength is 250 gausses and whose diameter is 10 cm. After deflection the entities strike a fluorescent screen which is located 20 cm from the center of the field. What distance will separate the traces made by the two types of particles?

4. What will be the mass of an electron when moving with a velocity of 2.7×10^{10} cm/sec.?

CHAPTER XXVI

ELECTRICAL CONDUCTION IN GASES

200. Discharge at Atmospheric Pressure. In Chap. VIII we studied the laws which govern the electronic current in **solids,** and in Chap. XII we dealt with conduction **in electrolytes.** In recent chapters we have considered the behavior of electrons in free space. It remains to examine the process by which conduction takes place in gases. The most widely accepted reference on this subject is a monumental two-volume work by the late Sir J. J. Thomson and his son, Professor G. P. Thomson. Various books by other workers, together with several thousand papers by those doing research in the field, constitute a substantial body of literature having to do with this subject. It is not an easy matter to condense the findings of so many investigators into a few pages. We will, however, endeavor to give a resumé of the theories and experimental findings connected with the movement of electrical charges through gases.

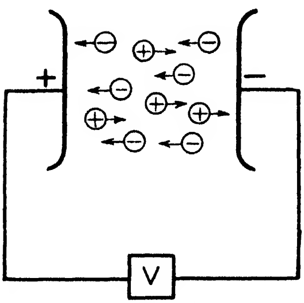


FIG. 303.—Discharge at atmospheric pressure.

To begin with, it may be set down as a fact that quiescent, dry, and dust-free gases show very little conductivity. However, as previously pointed out, it is probable that, at all times and under all circumstances, there are a few scattered free electrons, and corresponding positive ions, in any gas, due to ionizing agencies such as cosmic rays, local radioactivity, or some other form of radiation. That being the case, it follows that the establishment of a suitable potential gradient will give rise to a movement of such electrons and ions as may exist within the body of the gas. The situation might be represented as sketched in Fig. 303. The free electrons will move toward the positively charged terminal and the positive ions toward the negative electrode; and this constitutes one form of electric current—a form in which both positive and negative entities are involved. If the terminals are not too near together, and if the potential gradient is sufficient, both the electrons and the positive ions may acquire enough velocity to bring about ionization by collision, and thus the process becomes cumulative.

The relation between the ionization current and the applied potential,

in the case pictured in Fig. 303, has been shown to be as indicated by the graph appearing as Fig. 304.

It is to be noted that when the potential difference is small the space current follows Ohm's law, roughly; but when the potential difference has reached some particular value such as V_1, the potential difference may be increased considerably without causing any material increase in the space current. I, then, represents the **saturation current**—what few existing ions and free electrons that exist are being moved to the electrode as fast as they appear. When, however, the potential difference is raised to some particular value, such as V_2, the space current begins to rise rapidly; so rapidly, in fact, that a disruptive discharge may occur. For a gas at atmospheric pressure and a temperature of 25°C this critical potential

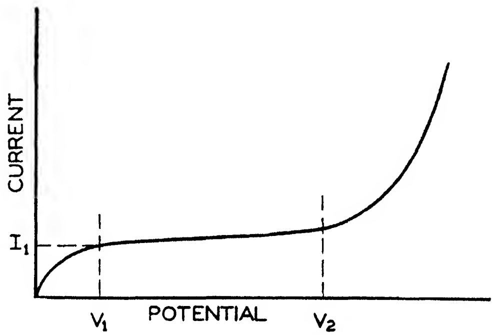


FIG. 304.—Relation between space current and the potential difference between the electrodes.

gradient is 30,000 volts/cm. If the spark-gap terminals are curved, the critical potential will be lower than the value given above, and if the electrodes are pointed, the breakdown potential will be still lower. The relatively low critical value, in the case of curved surfaces, is due to the fact that the field strength near the electrodes increases as the curvature increases; hence the potential gradient near such surfaces is comparatively high (Sec. 10). It has been found that the distance between curved electrodes is a factor in determining the critical potential in such cases. The breakdown potential gradient is found to **decrease** with increasing gap length. However, if the terminals are pointed, the critical potential gradient is fairly constant. It has been found by experiment that for distances up to several meters the critical gradient is of the order of 4,000 volts/cm. Owing to the factors which have a bearing on the results, it is extremely difficult to make accurate observations of discharge phenomena; hence the values given above are only rough approximations.

In passing, it should be noted that, once a disruptive discharge is initiated between two terminals which are connected to a source which is capable of supplying a substantial amount of energy, what is referred to as an **arc** may form. Because of the cumulative ionization by collision, and ionization due to ionic bombardment of the electrodes, the space between the terminals becomes almost **completely** ionized, with

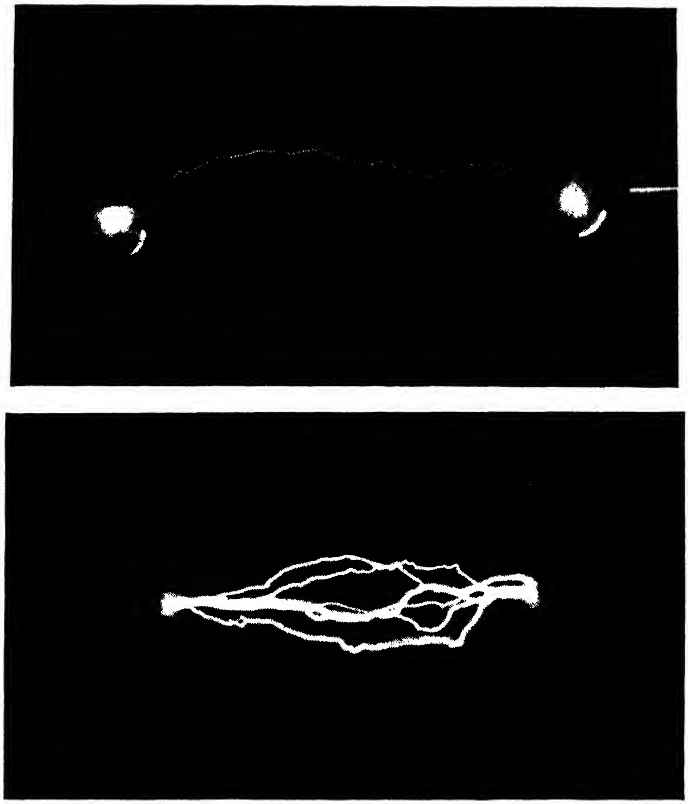


FIG. 305.—Photographs of disruptive discharges in air. The upper illustration shows a single discharge; while the lower picture shows a series of discharges. Note that the successive discharges do not follow a common path.

the result that an electronic and ionic current of substantial magnitude may result. This current strongly heats the electrodes, particularly the positive terminal, thus giving rise to additional electrons and positive ions. Unless the supply circuit includes a ballast resistor the terminals may be melted. In this connection it is also to be observed that once the arc forms only a relatively low potential difference is required for its maintenance and, as the arc current increases, less and less potential

difference is required. It is thus evident that **Ohm's law does not obtain under these circumstances.** Because of the current-voltage relation just mentioned, an arc is said to show a **negative resistance.** The potential difference required to maintain an arc discharge will depend, among other factors, upon the length of the arc and the nature of the electrode material. It requires but 50 volts to maintain a heavy arc between carbon electrodes in air. Two photographs of disruptive discharges are reproduced in Fig. 305.

Under certain circumstances, instead of a disruptive discharge between curved surfaces, what is known as a brush, or corona, discharge may occur. If the surfaces are relatively far apart the potential gradient may be sufficiently high (30,000 volts/cm or more) to produce rapid ionization in the **immediate vicinity** of the surface, but not great enough to produce marked ionization at some distance. In this event it is found that a conductor, such as a high potential wire for example, will be surrounded by an envelope of strongly ionized air. The result is that, due to electrical repulsion, a discharge from the conductor takes place. The energy thus dissipated constitutes a transmission loss.

201. Lightning. In the case of lightning we have a disruptive discharge on a scale involving great distance and correspondingly high potential differences. In considering this natural phenomenon we may first examine into the **cause** of such discharges.

First it may be set down as a fairly well-established fact that it does not have its origin in the atmosphere, in general. In this connection, it is true, as we have previously indicated, that a potential gradient exists in the atmosphere (100 $\pm$ volts/m); but it is interesting to note that the magnitude of this gradient is greater in winter than in summer. Further, the potential gradient existing between the surface of the earth and a point at some definite altitude is roughly the same for all parts of the world. These and various other important observations lead observers to believe that the electrical charges which go to make up the discharge we know as lightning have their origin in the clouds from which rain falls.

It is a well-known meteorological fact that a strong upward current of air exists near the front boundary of a moving cumulus cloud or "thunderhead." There is evidence for believing that this upward rush of air causes a mechanical disruption of raindrops that have attained a certain size, and that this breaking up of the drops causes ionization. As a result of this process the lower part of the cloud formation becomes strongly charged positively and the upper portion negatively. The situation is roughly represented by the sketch in Fig. 306. The potential gradient within the cloud itself is evidently great, with the result that lightning occurs more frequently between the two charged portions of the

same cloud than between any other two regions. Since the lower part of a storm cloud is positively charged a corresponding negative charge will be induced on the earth's surface below the cloud. If the potential gradient becomes great enough, ionization may build up rapidly and a discharge may occur between the lower part of the cloud and the earth. Occasionally discharges take place between two different clouds, though this occurs less frequently than the other two types of release.

The length of lightning discharges is probably of the order of 2 to 3 km (1.243 to 1.864 miles) when it takes place between a cloud and the earth, but the flash may reach up to 20 km (12.43 miles) when the breakdown occurs between two separate clouds.

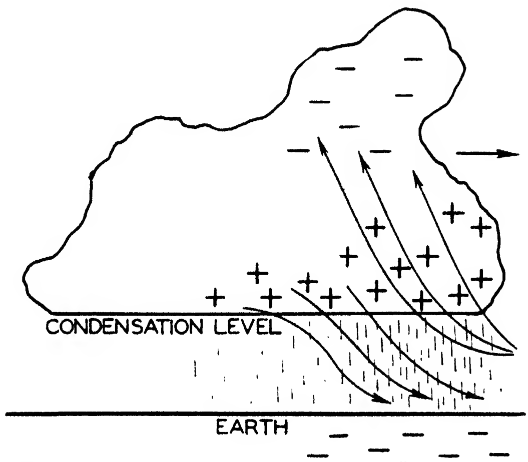


FIG. 306.—Showing how a potential difference is developed between different parts of a cloud and between a cloud and the earth.

The time involved in a **single** discharge is of the order of 0.001 sec but the complete discharge may occupy as much as 0.1 sec. It is difficult to arrive at any satisfactory estimate as to the magnitude of the current that passes in a lightning discharge. By indirect means however, it is, estimated that the current may reach a maximum value of perhaps 100,000 amp. On the basis of the above-indicated time factor, the quantity of electricity involved would therefore be of the order of 100 coulombs—not a great amount of electricity.

When it comes to the matter of protection, it may be said that it is clearly established that lightning rods, **when properly installed,** materially reduce the hazard from lightning. A lightning rod serves to alter the **direction** of the electric force near the surface of the earth. The conductors, which are functioning as lightning rods, should be well

grounded. This means that a connection **which will remain permanent** should be made to a metal plate so located in the ground that it is **at all times** in intimate contact with **moist** earth. A lightning rod system without an effective ground connection is a hazard—not a protection. The conductor forming the protective system **should not be insulated** from the building; and all gutters, water pipes and other similar metal bodies should be electrically connected, near their upper end, to the lightning conductors. It is recommended that the reader consult Professor Humphrey's interesting and informative volume, "Physics of the Air," particularly Chap. XV and XVI.

The protection of power and communication lines from lightning has been the object of extensive reasearch. Not only is damage done by direct hits, but marked line disturbances may occur due to inductive effects. Various forms of so-called lightning arresters have been developed which function more or less effectively. One widely used form employed for power-line protection consists of an assembly of semiconducting disks in series with one or more short air gaps. A high-potential surge in the line will cause the arrester to convey the line charge to earth, but when the line potential drops to its normal value the resistance of the arrester will not permit any current to flow to earth. Detailed descriptions of such protective devices are to be found in any standard work on power or communication engineering.

202. The Aurora. This display of nebulous light which sometimes appears in the northern portion of the sky in this hemisphere appears to be due to a luminous discharge caused by the ionization of the upper regions of the atmosphere by α-particles (Sec. 230) which have their origin in radioactive substances in the sun. Auroral displays are accompanied by magnetic storms, and the frequency of their occurrence is related to the sunspot period. Spectroscopic examination of the auroral light indicates that much of the radiation is due to nitrogen. There is also a prominent green line (5,578 Å) in the auroral spectrum, the origin of which is not known. The height of this ionized region above the earth varies from 85 to 1,100 km. The α-particles which bring about the ionization appear to be deflected by the earth's magnetic field, following the direction of the lines of force toward the magnetic pole.

203. The Ionosphere. It is recorded that Balfour Stewart was the first to suggest that an electrically conducting layer exists in the higher regions of the atmosphere. In 1902, A. E. Kennelly of Harvard suggested that the existence of ionized upper layers in the atmosphere might account for certain otherwise inexplicable radio transmission. A few months after Professor Kennelly's original suggestion was made, Oliver Heaviside of England independently brought forward a similar sugges-

tion. The existence of such an ionized region has been experimentally confirmed and its height determined. This ionized region extends from about 100 km to something like 200 km. At those altitudes the density of the atmosphere is very small, and since the relative composition of the air varies with height, there is a tendency for this ionized region to become stratified. There are two principal strata commonly recognized, the E and F layers, the former being the lower of the two. In connection with certain communication problems these two main strata are sometimes subdivided; but for our present discussion the E and F layers are all that need be considered. It is thought that the ionization of these upper atmospheric layers is brought about chiefly by ultraviolet radiation from the sun. Since this solar radiation varies, it might be expected that the height of the ionized layers, and also the degree of ionization, would be variable; and indeed experience shows such to be the case. When we come to study the propogation of electromagnetic waves we shall see how this ionized region plays an important part in radio communication. In that connection they function as conducting, and hence reflecting, surfaces. For a number of years this upper region of the atmosphere was referred to as the **Kennelly-Heaviside layer.** It is currently designated as the **ionosphere.**

204. Discharge at Low Gas Pressures. We have just considered several cases in nature which involve ionization, on a large scale, of gases at low pressure. We are next to examine certain phenomena which appear when ionization occurs in partially evacuated enclosures.

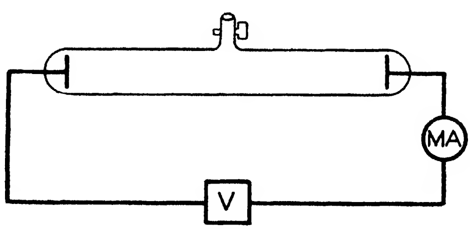


FIG. 307.—Arrangement whereby a discharge at low gas pressure may be studied.

Suppose that we have a tube, say 50 to 100 cm long, as shown in Fig. 307, and that a source of high potential difference is available, as indicated by V. We shall further assume that the tube is connected to an operating air pump to which a pressure gage is attached. The procedure will be to slowly evacuate the tube while a potential difference of perhaps 25,000 volts is applied to the sealed-in electrodes. At atmospheric pressure no discharge will take place. However, when the pressure has been reduced to a few centimeters of mercury it will be found that a wavy, colored

discharge appears between the electrodes, and the meter will indicate that current is flowing: ionization is taking place. If we reduce the pressure still further, say to a few millimeters, the discharge increases in volume until it fills the entire cross section of the tube. The ammeter will now show an increased reading, indicating that the resistance of the tube path has greatly decreased. When the evacuation progresses until the gage shows the pressure to be something like a few tenths of a millimeter a remarkable change occurs. The general luminous glow resolves itself into several fairly well-defined regions. In the immediate vicinity of the negative terminal (cathode) a limited colored region, called the **cathode glow,** is to be seen. Next to this is a narrow region showing little if any luminosity, and designated as the **Crookes dark space,** after the discoverer. Beyond the dark region just referred to, is to be seen a second and larger luminous section, which is designated as the **negative glow.** Still further along the tube we find a second and larger dark region, called the **Faraday dark space.** The remainder of the tube is filled with a more or less striated glow, ending in an anode glow at the positive terminal. The striated region is referred to as the **positive column.**

Before considering the final stage of our experiment it may be well to pause for a moment to consider the ionization process which has resulted in the phenomena thus far described.

As the pumping process progresses, ionization by electronic impact builds up rapidly. There are several reasons for this: (1) as the gas pressure decreases the mean free path of the electrons increases, and hence the kinetic energy of the free electrons tends to increase; (2) the chance of making impacts increases as the electronic velocity increases; (3) a hit may not detach an electron from an atom but it may result in exciting the atom, *i.e.*, one or more of the orbital electrons may acquire enough energy as a result of an impact to jump to a higher energy level. If this particular atom is not again hit within, say, 10^{-8} sec it will probably revert to its original condition; the electron which took up a larger orbit will drop back to its original energy level, and **the energy thus released will appear as radiant energy**—thus giving rise to the luminous effect which was observed. Now other hitherto unexcited atoms may absorb a portion of this radiant energy and, in turn, become excited, thus becoming an easy prey for swiftly moving electrons or positive ions. And so the process builds up. In considering any ionizing process it is to be noted that a moving electron must possess kinetic energy in excess of a certain **minimum** amount in order that it may ionize a normal atom. This minimum **energy** is referred to as the **minimum ionizing potential.** Each type of gas atom has its own particular ionization potential. Further, the radiant energy which may be absorbed by an atom, thus

causing it to become unstable, is commonly designated by the term **resonant radiation.**

Under the conditions that obtained in the second stage of our experimental discharge, the potential distribution within the tube between the

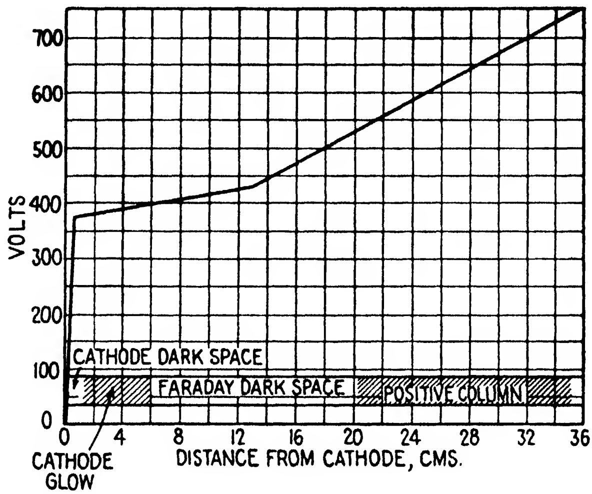


FIG. 308.—Potential distribution when a discharge is taking place at low gas pressure. (*Found and Forney, Trans. Am. Inst. Elec. Engrs.*, **47**, 4, 1928.)

electrodes would be as indicated by the graph shown as Fig. 308. It will be noted that the greatest potential drop occurs between the cathode and the negative glow region.

If we continue our evacuating process, the discharge enters what might be termed a third stage. When the pumping process has reduced the pressure to something like 0.01 mm the Crookes dark space increases in extent until it occupies the major part of the tube, the walls of the tube begin to fluoresce, and the current through the tube shows a **decrease**; thus indicating that the interelectrode resistance is **increasing.** If the evacuation is pushed still higher, less and less current flows; when the highest feasible vacuum is reached **the space current is entirely made up of electrons.** As this final stage is reached other and highly important phenomena present themselves; we shall deal with these in a later chapter.

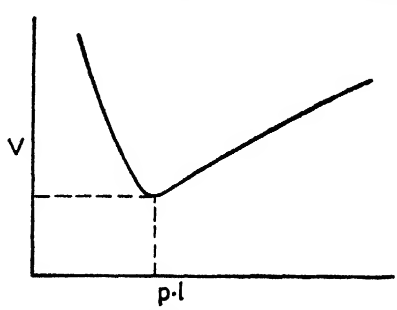


FIG. 309.—Illustrating the operation of Paschen's law.

205. Paschen's Law. In dealing with the passage of the current through a partial vacuum, the gas pressure, the electrode spacing, and the potential difference required are all interrelated. In 1889, Paschen found that the breakdown potential could be expressed as a function of the product pl, where p is the pressure of the gas and l the length of the discharge gap. This relation for a particular gas is shown graphically in Fig. 309. From an inspection of the curve it is evident that there is a certain **minimum** voltage, corresponding to a definite pressure and electrode separation, at which breakdown will occur. In a tube of uniform cross section the volume of the gas between the electrodes will vary as the product pl, and hence the **number** of atoms will vary in a corresponding manner. It would therefore seem that a certain definite number of atoms must be available in order to make a discharge possible at the minimum potential difference. The fact just mentioned is undoubtedly intimately related to the mean free path of the ionizing entities. It is to be observed that if and when pl has a value greater or less than the particular magnitude indicated, the corresponding breakdown voltage is greater.

206. Positive Rays. Before leaving the subject of the conduction of electricity through gases, one highly important aspect of the subject remains to be considered. In 1886, Goldstein, when working with a perforated cathode, observed luminous streamers of light behind the cathode. He assumed that this phenomenon was due to a stream of rays of some kind moving in a direction opposite to that of the cathode particles. Goldstein called these rays **kanalstrahlen** (canal rays, in English). By observing the effects of an electrostatic and a magnetic field, Wien, in 1898, established the fact that the entities constituting these so-called rays are in reality positively charged particles. He determined the velocity of the entities which compose this positive stream, and also the ratio of their charge to the mass (e/m). He found that the ratio e/m might have various values in the neighborhood of 10^4, which is of the same order of magnitude as in the case of the hydrogen ion in electrolysis. It was thus evident that the bodies that make up the positive stream have much greater mass than the electrons composing the cathode rays. In fact, the evidence appeared to indicate that their mass is about the same as that of an ordinary gas atom. It was found that these rays travel in straight lines, that they give rise to fluorescence, and that they cause the residual gas to emit radiation, but of a different color than that produced by cathode rays. It is to be noted that positive rays do not give rise to X rays.

In 1906, Sir J. J. Thomson began an extensive investigation of these **positive rays,** as they had come to be called. Since his procedure has

served as the basis of later methods of investigation, it will be worth while to review briefly the theory and technique of his experiments. Thomson worked with gases at relatively low pressures, and made use of

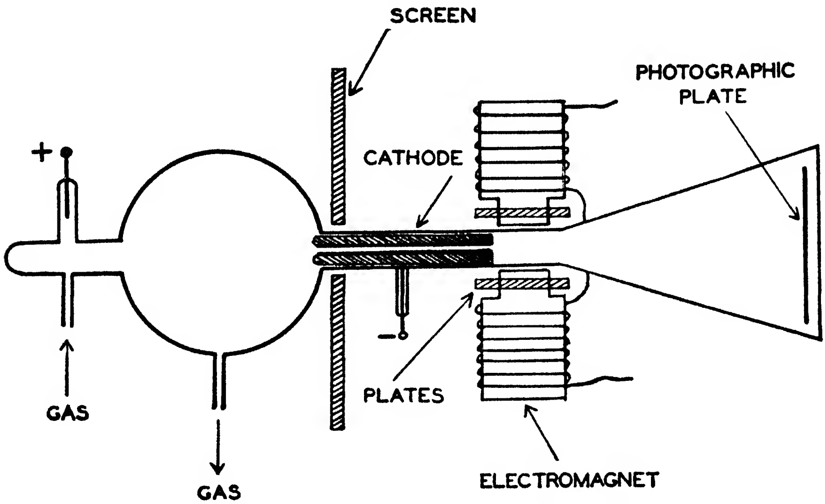


FIG. 310.—Assembly used by Sir. J. J. Thomson in his investigation of positive rays. The modern mass spectrograph was developed from this equipment.

a photographic method for recording the trace made by the rays. The object of his investigation was to determine accurately the e/m ratio. Having done that he could readily determine the magnitude of the masses involved.

The essentials of his experimental setup are shown in Fig. 310. Cold-cathode emission was utilized, the accelerated electrons passing through a very narrow collimating tube in the cathode. The positive-ray beam, which apparently originates just in front of the cathode (left end), was caused to pass through an electrostatic field due to the two charged plates, and also through a superimposed magnetic field established by the electromagnets shown. From the sketch it will be evident that the electrostatic flux and the magnetic flux are **parallel in direction.** The two resulting **forces,** however, are **not** parallel. Figure 311 indicates the relative direction of these forces and also the direction of the positive

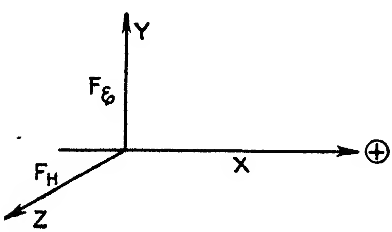


FIG. 311.—Relation of the electric and magnetic forces to the direction of the positive rays in the Thomson apparatus shown in Fig. 310.

rays. What will be the effect of these two forces on the entities making up the beam? Let us first consider the effect of the electric field.

The force acting on the charged bodies will, by Eq. (4), be given by

$$F_y = \mathcal{E}e = ma_y. \tag{i}$$

The deflection caused by this force will be

$$y = \tfrac{1}{2}a_y t^2 = \tfrac{1}{2}\frac{\mathcal{E}el^2}{mv^2}, \tag{ii}$$

where l is the length of the charged plate and v the velocity with which the body enters this composite field. Equation (ii) may be written

$$y = \frac{e}{mv^2}\beta, \tag{iii}$$

where β is a constant.

Following the same line of analytical attack as was utilized in Sec. 195, it may be set down that the force due to the magnetic field will be given by

$$F_z = Hev = ma_z. \tag{iv}$$

The deflection due to the magnetic field alone would accordingly be

$$z = \tfrac{1}{2}a_z t^2 = \tfrac{1}{2}\frac{Hel'}{mv}, \tag{v}$$

where l' is the length of the path through the magnetic field. The last equation may be given the form

$$z = \frac{e}{mv}\gamma, \tag{vi}$$

where γ is a constant.

A comparison of (iii) and (vi) shows that any displacement due to the electric field alone will vary inversely as the **square of the velocity;** while, in the case of the magnetic field, the effect will vary inversely as the **first power** of the velocity. An expression giving the displacement due to the combined effects of the two fields may be found by combining (iii) and (vi) through the elimination of the common factor v. Such a procedure leads to

$$z^2 = \kappa\frac{e}{m}y, \tag{264}$$

where κ is a constant incorporating β and γ. This, we see, is **the equation of a parabola.** Therefore if the ratio e/m has a fixed value, and the velocity of the individual particles vary (as would be the case under the circumstances), these entities would form a **parabolic trace** on the receiv-

ing screen or photographic plate. Since the charge e has a fixed value, a separate parabola would be formed by each group of positively charged entities having a common m value. Referring to Fig. 312, curve (1) would be made by particles having a smaller mass than those giving trace (2). For a given mass group, those having the smallest linear velocity would strike the receiving screen nearest the origin o, and those of high velocity, near the outer end. It is thus seen that this procedure should give a method whereby the velocity, the ratio e/m, and (knowing e) the mass of the particles forming the positive rays might be determined. Fortunately, the positive particles affect a photographic plate. Thomson

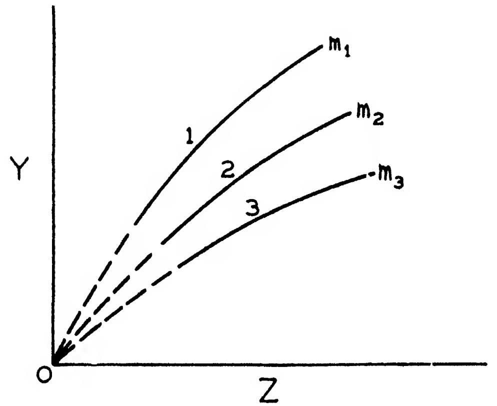


FIG. 312.—Parabolas formed by positive rays having different mass values.

introduced such a plate into the tube itself, as shown in Fig. 310. A reproduction of one of his photographic recordings is shown in Fig. 313. It is to be noted that several parabolas appear simultaneously, meaning that particles having different masses exist in the positive stream. By reversing the magnetic field during one half of the exposure, the negative half of the parabola can be secured, thus enabling the experimenter to determine the position of the Z-axis. This has been done in the case illustrated.

A study of the records indicated to Thomson that the entities which make up the positive rays are **positive ions,** *i.e.*, gas atoms which have lost one or more electrons, probably as a result of being struck by high speed electrons in the dark space just in front of the cathode.

By the method outlined above, Professor Thomson studied the positive ions of a number of atoms, including carbon, oxygen, mercury, and neon. In fact, he found the positive-ray method to be a highly sensitive means of chemical analysis, only an exceedingly small amount of material being required in order to carry out such an investigation. (The reader

is referred to Thomson's book, "Rays of Positive Electricity and their Application to Chemical Analysis.")

However, what is probably the most significant result of Thomson's research was his discovery that there are **two forms of the gas neon.** The most highly purified samples of neon available showed two parabolic traces, and measurements of these records indicated that neon may exist in two forms, one of which has an atomic weight of 20 and the other 22.

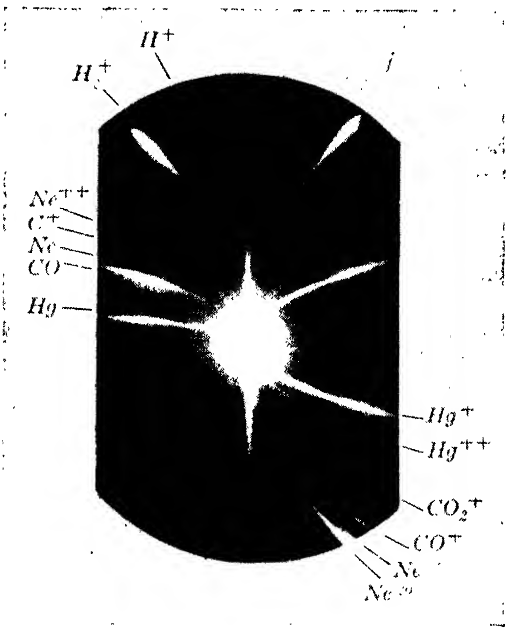


FIG. 313.—Photographic record of parabolas made by Sir. J. J. Thomson in his study of positive rays. Note the two curves corresponding to neon 20 and neon 22. This is an historical picture. (*From "Mass Spectra and Isotopes" by Dr. F. W. Aston. Reproduced by permission of Edward Arnold & Co., publishers.*)

Since the atomic weight of neon, as determined by chemical methods, is known to be 20.2, it is evident that what we commonly deal with is a mixture of these two forms of the same gas. Such different forms of a given element are called **isotopes.** The isotopes of an element have the same atomic number, but different atomic masses. In the case of neon, the greater part of the gas consists of the lighter form, as is evident from the Thomson photographic record (Fig. 313).

Thomson's tentative conclusions in the case of neon, and the wider implications of these findings, were presented before the Royal Institute in 1913. This discovery ranks as one of the most important in all

scientific history. Here we have disclosed the remarkable fact that an element may, and in most cases does, exist in a variety of forms. Thus the number of possible forms of the 90-odd chemical elements is enormously increased.

207. The Mass Spectrograph. In order to improve the accuracy of the observations originally made by Thomson, Dr. F. W. Aston, also of the Cavendish Laboratory, introduced a number of modifications in the positive ray setup originally used. The improvements made it possible to "focus" all ions of a given mass, even though their velocity may differ considerably. As a result, the resolving power of the apparatus was increased to the extent that one could distinguish between ionic masses

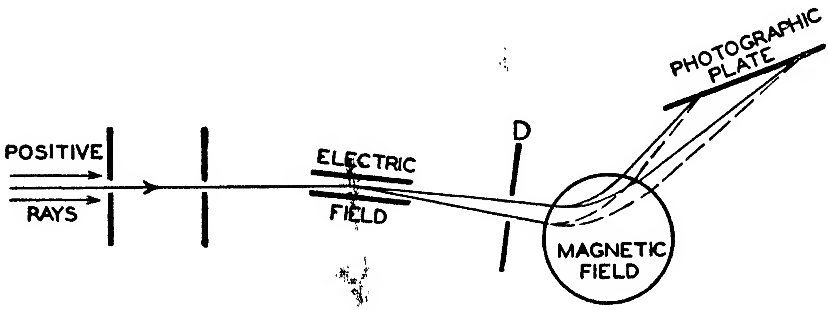


FIG. 314.—A mass spectrograph of the Thomson-Aston type. *(From "A Text-book of Physics" by C. A. Culver. Reproduced by permission of The Macmillan Co., publishers.)*

which differ by only 1 part in 600. The best that Thomson's equipment would do was 1 part in 20.

The essentials of the assembly by means of which this improved technique is carried out is sketched in Fig. 314.

In Sec. 206 we saw that the electric field, when acting alone, produces a deviation $\propto$ (e/mv^2); and also that the deviation due to the magnetic field is $\propto$ (e/mv). By passing a beam of positive rays first through an electric field and **later** through a magnetic field, Aston was able to effect a dispersion which depended **only** on e/m. A group of ions, then, having a certain mass, regardless of their velocities, were brought together at a common place on a photographic plate. Since dispersion in this apparatus is a function of mass, the assembly has come to be spoken of as a **mass spectrograph.** The analytical treatment of the problem was first given by Dr. Aston in a paper which appeared in the *Philosophical Magazine*, Vol. XXXVIII, p. 707, December, 1919. A good account of Aston's focusing method is also to be found in "The Particles of Modern Physics" by J. D. Stranathan.

Using the improved equipment, Aston confirmed Thomson's preliminary findings regarding neon. He also examined various other gaseous elements and found that they, too, exist in several forms. Demster, Brainbridge, and others have recently brought out improved models of mass spectrographs, with the result that a high degree of resolution is currently possible. All of the elements have now been studied with regard to possible isotopes, and complete tables are to be found in any work on chemistry or atomic physics. The mass spectrograph has now become an engineering tool that is used in industrial laboratories. The student may well read "Mass Spectra and Isotopes" by Aston.

208. Geiger-Muller Counter. Before closing the chapter on gaseous conduction, it will be worth while to examine the structure and function-

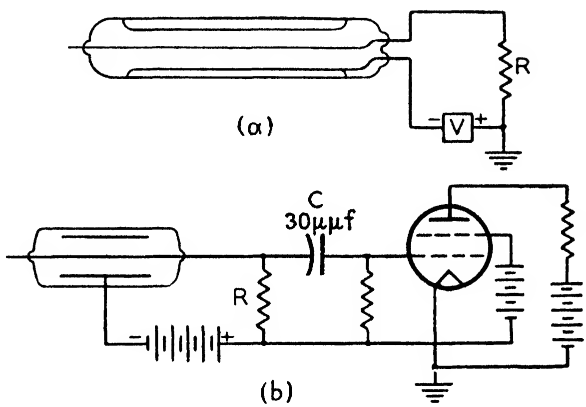


FIG. 315.—Geiger-Muller counter assembly.

ing of a widely used device which depends for its operation on the intermittent ionization of an attenuated gas.

As previously indicated, various agencies may operate to ionize a gas, among which may be mentioned, in addition to ultraviolet radiation, X rays, gamma rays, α-particles, and cosmic rays, all of which will be dealt with in detail in later chapters. The Geiger-Muller counter is a two-electrode gas-filled tube which may be used for the purpose of detecting and measuring corpuscular and quantum radiation by means of the ionization which they produce. This device is made in a number of different forms; a common type is shown diagrammatically in Fig. 315*a*. This form consists of a thin-walled copper tube approximately 3 cm long and 1 cm inside diameter. This is supported within a sealed glass housing. A tungsten wire, 0.08 mm in diameter, welded to heavier seal-in wires, is supported along the long diameter of the tube. The copper tube is connected to the negative terminal of a constant high-

potential source V, thus becoming the cathode; the centrally located wire is connected to the positive lead, through a resistance R, of the order of 500 megohms, thus becoming the anode. The positive terminal of the potential source is grounded. The tube is filled with a mixture of 94 per cent argon and 6 per cent oxygen at a pressure of 5 to 10 cm. The voltage of the d-c source V is slightly less than that which will cause a discharge between the electrodes when ionization does not obtain. With this particular unit the starting potential (potential at which a response can be obtained) is 698 volts, and the operating potential is 780 to 800 volts. Counters of this character may be designed to be sensitive to α-, β-, and γ-rays; they will also respond to X rays and cosmic rays. If and when an electrically charged entity, or a quantum of radiation, passes through the glass and copper walls and into the region between the cathode and anode, ionization of the enclosed gas will take place. The resulting cumulative ionization will give rise to a relatively large pulse of current through the tube and its associated circuit. This rush of current will result in a pronounced drop in potential across the series resistor R, which in turn will cause the voltage across the tube to fall below the starting potential, thus interrupting the space current; the tube is then ready to respond to subsequent excitation. The deionization time is very short, with the result that the tube will respond to ionizing impulses at a rapid rate. The current in such a circuit is of the order of 10^{-8} amp. With a series resistor of the magnitude indicated above, the drop would be about 5 volts. The drop across this series resistor is usually impressed, through a low-capacitance condenser of high resistance, on the grid of the first tube of a multistage amplifier. The output of the amplifier may be arranged to operate a meter, a counter, or a loud-speaker. The diagram of Fig. 315*b* shows the essentials of a counter-amplifier circuit.

The Geiger-Muller counter has found wide use in research work connected with nuclear physics and cosmic rays.

CHAPTER XXVII

ROENTGEN OR X RAYS

209. Discovery of X Rays. The closing years of the nineteenth century were rich in scientific discovery. We have already noted the identification of the electron by J. J. Thomson in 1897. Later, we shall have occasion to examine into the nature and result of another discovery of profound importance, which occurred in 1896. In this chapter we are to consider a third epochal contribution to the progress of physical science.

In the closing months of the year 1895, Dr. Wilhelm Konrad Röntgen, professor of physics at Würzburg, Bavaria, began a series of experiments with cathode rays. The records show that he was evidently looking for new phenomena when he made the discovery which bears his name. The fact that he had a fluorescent screen at hand, and that he had completely covered with light-proof paper the evacuated tube which he was operating; indicates he was studying the possibility of some form of dark (invisible) radiation. True, the radiation-sensitive screen was lying on the table instead of being held in the hand, but that does not make the discovery accidental. Earlier, others had observed similar phenomena, among them Professor Goodspeed at the University of Pennsylvania; but because of his highly developed experimental skill and keen powers of observation Röntgen immediately grasped the significance of the fluorescence exhibited by the screen. Thus in November, 1895, an epoch-making discovery was made. Röntgen immediately followed up his preliminary observations by carrying out a comprehensive research on the properties of this newly discovered form of radiation, and in December he presented a paper embodying the result of his investigations. Röntgen established that X rays

1. Show rectilinear propagation
2. Cause fluorescence when incident on certain substances
3. Affect a photographic plate
4. Pass through many substances commonly opaque to visible and ultra-violet light
5. Are absorbed by matter
6. Ionize gases
7. Are not identical with cathode rays
8. Originate where and when cathode rays impinge on matter
9. Are not affected by electric or magnetic fields

With the means at hand, he was unable to detect reflection, refraction, interference, or polarization of the rays. Later investigators have observed all four of these phenomena, as we shall see presently. Not knowing the exact nature of the new type of radiation he had discovered, but having observed that they produced definite shadows, Röntgen gave them the tentative designation of X rays.

210. Production of X Rays. Röntgen was using a Hittorf-Crookes tube when he made his original observations. The news of his discovery spread with great rapidity, and soon special tubes were designed for the generation of these rays. The most common early model had the form indicated in Fig. 316. A concave cathode served to focus the cathode

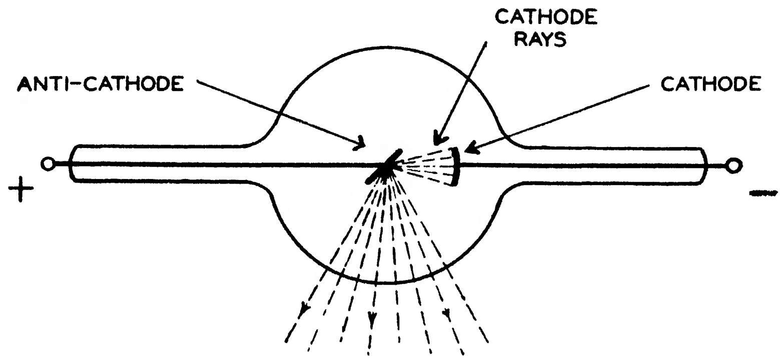


FIG. 316.—Early form of X-ray tube.

emission on a target, which also functioned as the anode. High-field emission was brought about by the application of a direct potential difference of the order of 30 to 100 kv. The gas pressure in the tube was of the order of 0.001 mm. The electronic bombardment of the target, or anticathode, gave rise to the X radiation. Much of the energy of the impinging electrons is transformed into heat; hence in order to avoid vaporization of the target material, it was necessary to use a metal such as platinum or tungsten as an electron-receiving surface. Because of the fact that the vacuum was a variable quantity, these early tubes were more or less unsatisfactory. In 1913, Dr. W. D. Coolidge of the General Electric Research Laboratory, encouraged by the work of his colleague, Dr. Langmuir, developed[1] an X-ray tube in which the space current is wholly electronic—thermionic emission is utilized. Because of its superior performance, the Coolidge type of tube is now universally used. It is constructed in a number of forms depending upon the use to which

[1] COOLIDGE, W. D., *Phys. Rev.*, **2**, 409 (1913).

the unit is to be put. A recent model of the Coolidge type of tube is shown in Fig. 317. Note the cooling vanes. Since the current through the Coolidge tube is electronic, the unit is self-rectifying and can therefore be operated directly from the secondary of an a-c transformer. The tubes commonly used in medical work operate at about 30 kv with a current of 10 to 50 ma, but tubes have been constructed which may be operated at several hundred thousand volts and with a power rating of 10 kw. Indeed, tubes are now in use for both medical and industrial purposes which are operated at a potential difference of 1 million volts.

In the operation of any X-ray tube the penetrating power of the rays produced depends upon the speed with which the electrons strike the target, which in turn is determined by the potential difference between the tube terminals; the higher the operating potential the greater the pene-

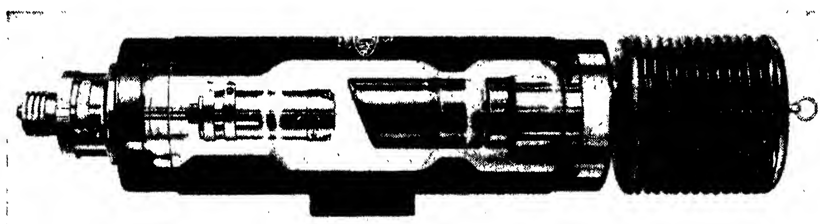


FIG. 317.—Phantom view of a modern X-ray tube. The outer envelope is of metal except for a glass window through which the rays emerge. The metal vanes on the right end serve to radiate the heat developed by the electronic bombardment of the target. (*Courtesy of Machlett Laboratories.*)

trating power. The kind of material constituting the target determines, to some extent, the character of the X-ray output; metals of high atomic number yield more intense radiation. With the Coolidge tube it is possible to secure a yield of intense homogeneous X radiation of any desired penetrating power. The current strength for the filament (cathode) determines the **intensity** of the rays emitted, and **the voltage applied to the tube determines the penetrating power,** or "hardness." These two factors are **not** interdependent; both are independently under the control of the operator. Because of the relatively large current-carrying capacity of the modern tube and the use of intensifying screens, it is now possible to make instantaneous skiagraphs of the vital organs of the human body. An intensifying screen consists of a fluorescent surface made of such a material as tungstate of calcium. Such a screen fluoresces with a highly actinic bluish light when subjected to X rays and, in use, is placed in contact with the film of the photographic plate. By the use of such a screen, the exposure necessary to secure satisfactory

skiagraphs may be reduced to a small fraction of the time formerly required.

211. Nature of X Rays. Thus far in our discussion the question of the nature of Roentgen rays has purposely been held in abeyance. It is now in order to raise this interesting and important query and to examine the evidence leading to modern conclusions about the character of this form of radiation.

Following the discovery of the new form of radiation, various suggestions were advanced as to the nature of the dark Roentgen rays. We shall not review these in detail, but only note in passing that the tentative theory which was given most serious consideration at the time was that due to Stokes, which was to the effect that X rays consist of

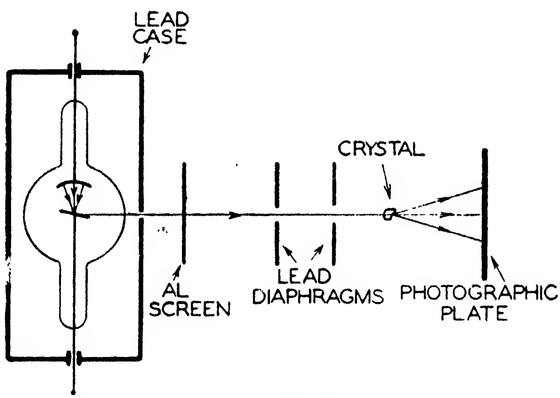


FIG. 318.—Assembly used in making Laue diffraction photographs.

irregular pulses in the ether. Owing to the fact that the early investigators had been unable to reflect or refract X rays, it was extremely difficult to arrive at a satisfactory explanation of their nature.

Proceeding on the assumption that Roentgen rays were similar to light waves but of extremely short wave length, Professor Max Laue,[1] of the University of Zurich, suggested that the regular arrangement of the atoms in a crystal might possibly serve as a "grating" for use in producing interference effects with X rays and thus yield data from which it would be possible to determine accurately the magnitude of the wave length. According to this suggestion the spacings between the atoms in the crystal were to serve as the grating spaces. Dr. Laue investigated the problem mathematically and formulated laws which pointed to the fact that one should be able to secure an interference pattern. Knipping and Friedrich[2] in 1912 carried out the necessary experiments to test

[1] M. LAUE, *Physik. Zeit.*, **14,** 421 (1913).

[2] FRIEDRICH, W., P. KNIPPING, and M. LAUE, *Le radium*, **10,** 47 (1913).

Laue's analytical results. These experiments were brilliantly successful and clearly confirmed Laue's predictions; thus establishing the fact that X rays are in no respect different from visible light except in the magnitude of their wave length. These and other experiments show that the wave length of X rays lies between 0.25 Å and 1.3 Å. The arrangement of apparatus necessary for the production of a Laue diffraction photograph is indicated in Fig. 318. Figure 319 is a reproduction of a typical Laue diffraction record. The central spot is due to the rays transmitted directly through the crystal, while the spots surrounding this central area are due to the interference caused by the atoms and interatomic spaces acting as a transmission grating. It was W. L. Bragg who pointed out that each spot in the pattern represents the diffraction of the incident X rays by a certain plane of the crystal structure which contains large numbers of atoms. Here we have a striking illustration of the experimental confirmation of pioneering analytical research. The investigation of Laue and his coworkers has led to far-reaching practical results. Once knowing the wave length of the radiation being used, one may utilize the Laue procedure to determine the structural arrangement of the atoms which go to make up crystalline bodies. In fact, the method is also applicable to noncrystalline substances. The making of Laue diffraction photographs for the purpose of studying the molecular structure of various manufactured products is now widely utilized in the industrial arts.

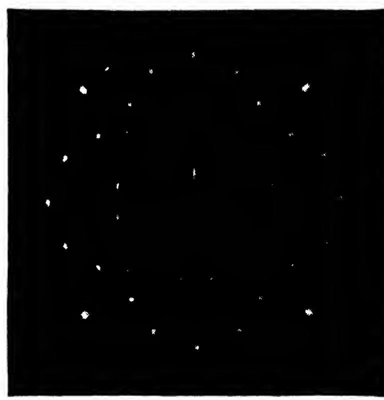


FIG. 319.—Diffraction photograph made by passing a beam of X rays through a crystal of sodium chloride.

212. The X-ray Spectrometer. In utilizing X rays by the Laue method in the determination of crystalline and other structures, it is found that the diffraction patterns are often highly complex, and hence difficult to interpret. Early in his work on X rays and crystal structure, Bragg suggested using a crystal as a reflection grating instead of as a transmission grating. Bragg's theoretical approach to the problem was somewhat like the explanation commonly given to account for the color of thin films in the study of light. As outlined by Bragg, the case is somewhat as follows.

In Fig. 320 the dots represent the atomic arrangement in a representative crystal. The distance apart d is, in general, of the order of 10^{-8} cm. R_1 and R_2 indicate two typical X-ray wave trains incident on the crystal

as shown. Diffuse reflection from the several atoms will occur. Let us consider that part of the several wave trains which is reflected from successive crystal planes at an angle θ, the reflected rays being designated as R_1' and R_2'. AB is parallel to the wave front of the incident ray, and AD is parallel to the reflected wave trains. Because d is extremely small, R_1 and R_2, as well as R_1' and R_2', will be very close together; consequently, so far as any record on a photographic plate is concerned, the reflected rays (if in phase) will produce a single impression. In reaching the plate, R_2 will travel a distance $(BC + CD)$ greater than R_1. If that difference in path is some **whole** number of wave lengths, the two wave trains will

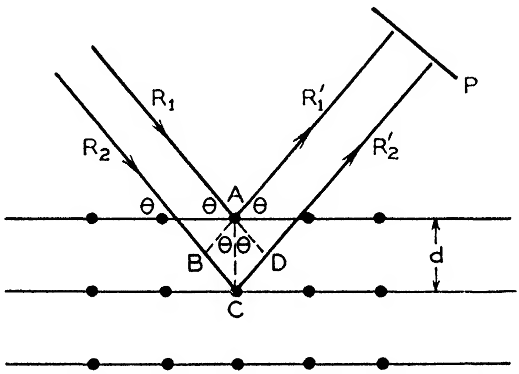


FIG. 320.—Showing the manner in which the atomic structure of a crystal may bring about the diffraction of X rays.

reinforce one another at P, and thereby give rise to a record on the plate. Stated algebraically, the situation would give

$$BC + CD = n\lambda.$$

From the geometry of the case

$$BC = d \sin \theta,$$

and

$$CD = d \sin \theta;$$

therefore

$$n\lambda = 2d \sin \theta. \quad (265)$$

The above expression is known as **Bragg's equation**. If and when $n = 1$, the difference in path for waves reflected from any two adjacent planes is one wave length; in which case the above relation becomes

$$\lambda = 2d \sin \theta. \quad (266)$$

If, then, we cause an X-ray beam to strike the face of a crystal and then measure the angle of the reflected beam, by photographic or other means, we can readily determine the wave length of the incident radiation. If the radiation is not monochromatic it will be found that other spots will appear on our recording plate corresponding to other angles of reflection. Thus the spectrum of the radiation may be completely determined. This is, of course, based on the assumption the atomic spacing is known. If the wave length is known, the atomic spacing can be determined by observing the diffraction patterns as one allows monochromatic X radiation to fall on the several faces of the crystal.

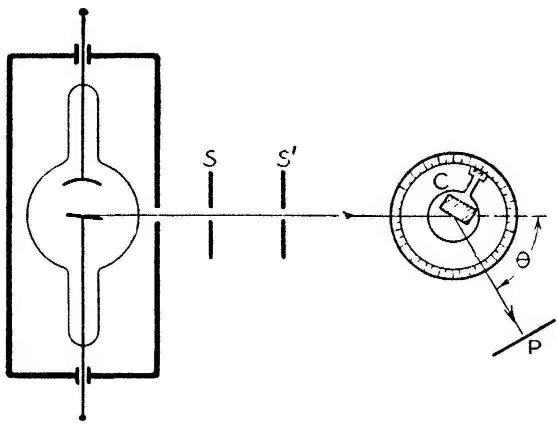


Fig. 321.—Arrangement for using an X-ray spectrometer.

A **spectrometer** for use with X rays, similar to the one used for visible radiation, can be readily assembled. The general mechanical relation of the several components is depicted in Fig. 321. A narrow parallel beam of X rays is secured by means of the collimating slits S and S'. The radiation is allowed to strike the face of a crystal (mounted on the spectrometer table) and the direction of the reflected beam is observed by means of a fluorescent screen or a movable ionization chamber, of the Geiger counter type, located at P. The angle that the reflected beam makes with the direction of the incident ray can then be read from the graduated circle on the instrument. By substituting a photographic plate for the ionization chamber, a record can be made, simultaneously, of any and all lines in the X-ray spectrum. The assembly then becomes an **X-ray spectroscope.** The crystals most commonly used in X-ray spectroscopy are rock salt, calcite, and quartz. By means of such a piece of equipment, and the application of Eq. (266), either the wave length or the crystal spacing may be determined.

213. The X-ray Spectrum. By means of the X-ray spectrometer it has been found that when high speed electrons are abruptly stopped by a target two types of X radiation appear. One part consists of a relatively wide band of frequencies, the intensity and limits of which depend chiefly on the magnitude of the accelerating potential. Superimposed on this more or less continuous spectrum there is also to be found a well-

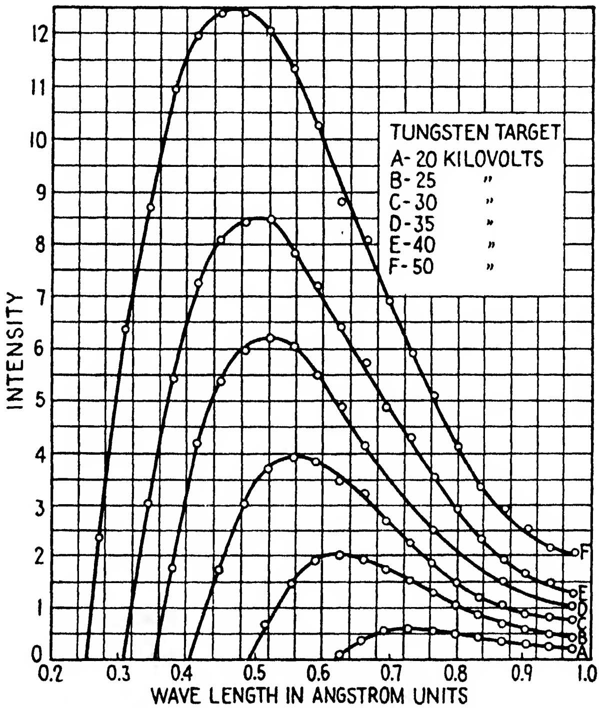


FIG. 322.—Showing the effect of increasing the anode potential on an X-ray spectrum. (*Courtesy of Physical Review.*)

defined line spectrum consisting of two groups of lines, one known as the K series and the other as the L series.

In the generation of X rays we are dealing with the inverse of photoelectric emission. In the case of X radiation the stopping of electrons gives rise to the emission of photons; in photoelectric emission photons act to liberate electrons.

Several aspects of the continuous spectrum are of special interest. Figure 322 is a reproduction of the heterogeneous X radiation originating at a tungsten target when the exciting electrons have fallen through

various potential differences.[1] It will be noted that, for a given exciting potential difference, there is a definite upper-frequency limit, and that this limit **shifts in the direction of the higher frequencies** as the accelerating potential is raised. Further, we see that the wave length having the maximum energy content shifts toward the shorter wave lengths as the potential is increased. (This is somewhat analogous to Wein's displace-

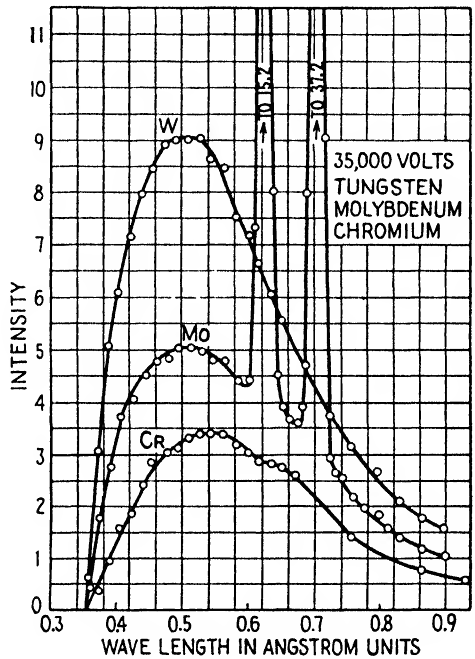


FIG. 323.—Showing the X-ray emission when different metals are used as target materials. Note the pronounced intensity maxima in the case of molybdenum. (*Courtesy of Physical Review.*)

ment law in the field of thermal and optical radiation.) In Fig. 323 the intensity-wave-length curves are shown for three different target materials, **the accelerating potential being held constant.** These graphs clearly show that **the upper frequency limit is independent of the nature of the target.** This aspect of X-ray production has been carefully studied (1915) by Duane and Hunt. These investigators found that the short-wave limit above referred to varies inversely as the potential difference across the tube. In mathematical form their findings would be expressed

[1] The graphs shown in Figs. 322 and 323 are reproduced from a paper by C. T. Ulrey, *Phys. Rev.*, **11,** 405 (1918).

as

$$Ve = h\nu_{\min} = \frac{hc}{\lambda_{\min}}, \tag{267}$$

where V is the accelerating potential difference, e the electronic charge, h Planck's constant, and c the velocity of light. The product Ve, of course, represents the energy with which an electron strikes the target. The above expression is known as the **Duane and Hunt law.** From it we see that, knowing V, e, and $\lambda_{\min}$, it becomes possible to determine the value of Planck's constant. Various investigators have computed the value of h on such a basis, and have found values the mean of which correspond closely with the value found by other methods.

214. Characteristic X Rays. Turning now to the line spectrum of X radiation, we find that Barkla and Sadler in 1908 made the highly important discovery that many substances (probably all) may be caused to emit a form of radiation which is characteristic of that particular substance, just as the spectral lines in the visible range are characteristic of the particular incandescent substance. They found that this characteristic X radiation could be produced by bombarding the substance with high-speed electrons, or by irradiating the substance with primary X rays. Since then it has been found that this characteristic radiation may be produced when a substance is bombarded by α-particles or by protons. When produced by electronic bombardment, characteristic X rays appear only when the electrons possess a definite minimum energy content.

Referring to Fig. 323, it is to be noted that the two lines of the K series for molybdenum are present, the exciting potential being 35 kv. At that particular voltage the characteristic lines for tungsten and chromium do not exist. If, however, the accelerating potential were raised to about 70 kv, the characteristic lines for tungsten would appear. It is found that as the atomic weight increases, correspondingly higher voltages are required in order to produce the characteristic spectra. It is known that an electron must fall through a potential difference of 12,260 volts in order to give rise to a wave whose length is 1 Å. We have just seen that nearly six times this value is required to produce the characteristic lines of tungsten, for instance. This means that the quanta emitted are relatively large, and the resulting frequencies are high. Hence, in such a case, the resulting rays are "hard." In any given case the radiation of the K series is, perhaps, 300 times more penetrating than the L emission.

Shortly before his untimely death in the First World War, Moseley of the Cavendish Laboratory at Cambridge University, utilizing the

X-ray spectrometer and the photographic method, carried out a comprehensive and important study of the line spectra of some 38 different elements.[1] In Fig. 324 is to be seen a reproduction of several of Moseley's photographic records, showing the *K* lines. It will be noted that these spectrograms are arranged vertically in the order of their atomic numbers (smaller numbers at the top). The record of any particular element is so placed that its position from the left side of the figure is roughly proportional to the wave length of the spectral lines. Several facts are at once apparent. (1) The frequency of the characteristic *K* radiation increases with the atomic number. (2) Except as to wave lengths, **each element exhibits the same type of spectrum.** (3) One could predict that there must be an element between calcium and titanium; we of course know that scandium occupies that position. (4) The wave having the longer wave length has the

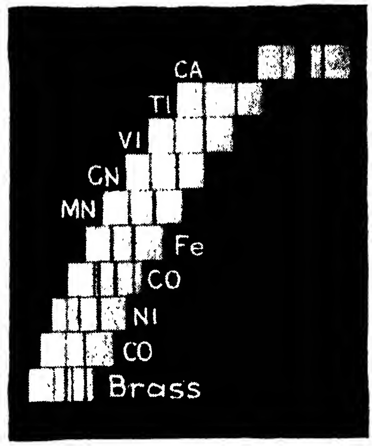


FIG. 324.—Photographs of X-ray spectra, *K* series, by Moseley. (*Courtesy of Philosophical Magazine and Journal of Science.*)

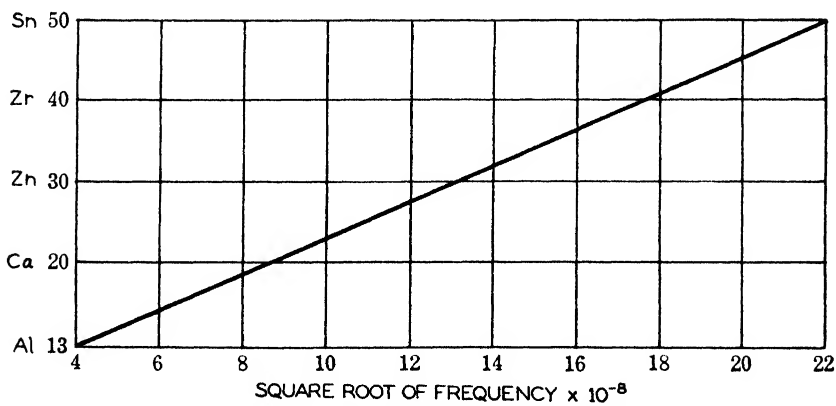


FIG. 325.—Relation between the atomic numbers of the target materials and the frequency of the X-ray emission.

greater energy content. (5) The alloy brass shows the characteristic lines of the constituent elements, *i.e.*, copper and zinc.

[1] MOSELEY, H. G. J., *Phil. Mag.*, **26**, 1024 (1913); **27**, 703 (1914).

When Moseley plotted the square root of the frequency against the atomic numbers he found the relation to be linear, as shown by the graph in Fig. 325. This is perhaps the most significant aspect of Moseley's work. It means that the atomic numbers are of the utmost significance. Indeed, the application of Moseley's law has made it possible to determine the proper relative position of several elements in the periodic table.

Other investigators have carried forward the work begun by Moseley, with the result that all of the elements from which the characteristic radiation can be produced have been studied. It has been found that the wave lengths of the lines of the L series also vary regularly with the atomic numbers in the same general manner as do those of the K group.

FIG. 326.—Spectrograms of the L radiation from several of the elements whose atomic numbers, reading from the top, are 79, 81, 82, and 83. (*After Siegbahn.*)

In Fig. 326 is to be seen the spectrograms of the L radiation from several of the elements. These particular records are due to Siegbahn.

In the application of Roentgen rays in the fields of medicine, surgery, and industry, the **continuous X-ray spectrum** is more important than the **characteristic radiation,** for the simple reason that by far the greater amount of energy is to be found in that part of the emission. However, in the domain of research the characteristic radiation has served as an important tool.

In passing, it may be noted that there is reason to believe that the electrons in the outer orbits (valence electrons) do not take part in the production of X rays. Evidence seems to point to the fact that those electrons nearest the nucleus are the ones that take part in the genesis of X rays.

215. Scattering of X Rays and the Compton Effect. When X rays are caused to be incident on matter, that material becomes a secondary source of X radiation and of secondary high-speed electrons. The intensity of the secondary X radiation is small compared with the primary, or incident, radiation. However, in the practical application of Roentgen rays, this secondary radiation must be considered and steps taken to

eliminate its effects. To the scientist, these secondary X rays are of considerable theoretical importance.

It has been found that, if a number of substances are irradiated by primary X rays, two definite types of secondary emission appear. One of these, known as **fluorescent X rays,** consists of radiation, the frequency of which depends upon the **nature of the radiator** only. This frequency is always lower than that of the incident radiation, in conformity with Stokes's law. The situation is essentially the same as the fluorescent phenomena one encounters in the case of visible radiation; hence the name. The fluorescent rays result from the excitation of the atoms of the radiator by the primary rays, the K and L types of radiation (Sec. 214) being produced.

The second type of secondary radiation, referred to as **scattered X rays,** is the result of a process which is somewhat like that involved in the scattering of visible radiation. On the basis of the quantum theory it may be supposed that certain photons may pass near, but do not strike, a given electron. During this near approach the photon may be deflected from its original course, its direction thus being changed, but its frequency remaining unaltered. If, however, a photon does make contact with an electron, the electron will acquire a part of the energy of the photon; and, on the assumption that the masses of the electron and photon are comparable, the electron will recoil from the impact. Thus the impacting photon will have less energy after collision than before. But if the product $h\nu$ is to change in value, it must be the frequency factor that changes. Hence the deflected photon may be expected to represent a lower frequency. This would mean that some of the scattered X rays would have a lower frequency (longer wave length) than the incident radiation, and such turns out to be the case.

In 1922, A. H. Compton, by means of an X-ray spectrometer, examined the scattered radiation from a radiator and found that the emission consisted of **two** frequencies, one of which was identical with the primary radiation and the other less than that of the incident rays; the latter composing the greater part of the scattered rays. The phenomenon is now known as the **Compton effect.** A complete discussion of the Compton effect involves the principles of quantum mechanics, and would take us too far afield. It must suffice to point out that the explanation involves the assumptions that quanta have the speed of light, that the photon energy $h\nu$ is kinetic, and that the well-known laws of the conservation of energy and momentum hold in the case just outlined. On those assumptions it is possible to develop the relation

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\theta) = 0.024(1 - \cos\theta), \tag{268}$$

where $\Delta\lambda$ is the increase in wave length, m the mass of the electron, c the velocity of light, and θ the angle between the primary rays and the scattered radiation. Graphically the situation may be represented as shown in Fig. 327. From the above equation it is evident that the observed change in frequency will depend upon the angle at which the observation is made. The experimental results secured by Compton and others confirm the foregoing relation. Even the recoil electrons have been observed. Thus we see that, in this case at least, it is justifiable to quantize X radiation. When we were considering X-ray diffraction it was convenient to do so on the assumption that we were dealing with

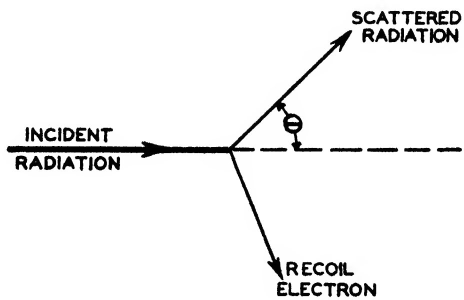


FIG. 327.—Compton effect.

a wave phenomenon. It would therefore appear that X radiation is dualistic in character.

216. Polarization and Refraction of X Rays. In 1905 Barkla demonstrated that primary X rays are at least partially polarized. On the basis of the classical electromagnetic theory of radiation this was to be expected. Barkla also found that scattered X rays are polarized, and to a greater extent than the primary beam. In this case, maximum polarization occurs for those rays which are scattered at an angle of 90°. The secondary radiation that proceeds directly forward or backward is not polarized.

At the beginning of our discussion of Roentgen rays it was indicated that early efforts to refract such rays were unsuccessful. However, in more recent years it has been found possible to bring about the refracton of this form of radiation. In 1919, Stenström observed that Bragg's law [(Eq. (265)] when applied to reflected monochromatic radiation did not give exactly the same value for θ in succeeding orders. He attributed this to the fact that at least some of the incident radiation entered the reflecting crystal and was slightly refracted before again emerging. This interpretation has been confirmed by the results secured by other workers. It has now been definitely established that X rays may be

refracted on entering a medium, but that the refracted ray is bent **away from the normal.** This means that the refraction index for X rays is **less than unity.** This is, of course, contrary to the behavior of light waves and the phenomenon raises some extremely interesting questions which we cannot pursue here. In passing, however, it is to be noted that the phenomenon of total reflection of X rays has been observed, as in the corresponding case in the realm of visible radiation.

217. The Betatron—A New X-ray Generator. Recently there has been developed, largely as a result of the research work of Dr. D. W.

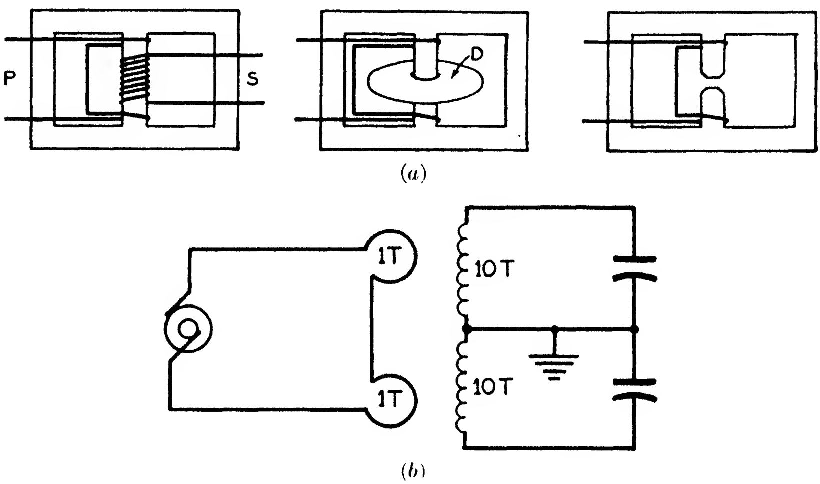


FIG. 328.—Fundamental circuits of the betatron. (*a*) Magnetic circuit; (*b*) exciting circuit. (*Courtesy of Electronics.*)

Kerst[1] at the University of Illinois, a piece of equipment by means of which it is possible to give electrons an unusually high linear speed. These high-speed electrons can be caused to produce copious and extremely penetrating X rays.

The principles involved in the operation of the new apparatus are simple. We have seen that a magnetic flux that changes with time will give rise to an emf; and it is to be noted that such an emf is produced regardless of the presence or absence of a conducting medium. Any free electrons in the region of a changing flux, whether they form a part of a conductor or exist in free space, will be acted upon by such an induced emf, and accordingly will undergo acceleration. This is the principle

[1] Dr. Max Steenback of Germany claims to be the inventor of this new device. However, credit for the development of a successful unit should unquestionably go to Dr. Kerst.

upon which the Kerst equipment operates, and is the reason why the device is sometimes referred to as an **induction accelerator.**

Basically, the **betatron,** as this new accelerator has come to be called, is essentially a special transformer—a transformer in which a doughnut-shaped evacuated chamber constitutes a single-turn secondary. The mechanical and electrical relations involved are sketched in Fig. 328*a*. In the diagrams, (1) shows an ordinary shell-type transformer; in (2) the usual secondary is replaced by a highly evacuated acceleration chamber; (3) shows the cross section of the magnetic circuit employed. A changing magnetic field is produced by an alternating current in the

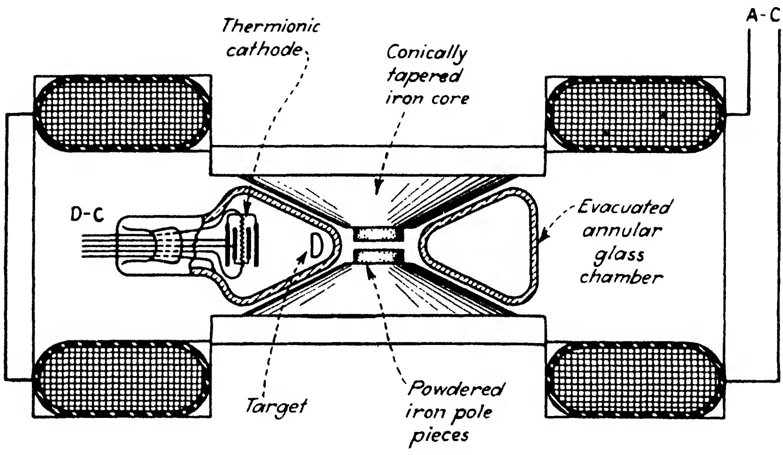


FIG. 329.—Diagrammatic sketch of the first model of the betatron designed by Kerst. (*Courtesy of Electronics.*)

circuit assembly indicated in Fig. 328*b*, the capacitor bank being inserted for the purpose of correcting the power factor in the transformer circuits.

Figure 329 shows diagrammatically a cross section of one of the earlier models designed by Kerst. Thermally produced electrons are introduced tangentially into the evacuated chamber by the periodic application of a high positive potential to an accelerating electrode. Each pulse of electrons is synchronized to occur at the beginning of each cycle of the magnetic field, as indicated by the points A and A' in Fig. 330. The initial accelerating potential is applied for only a few microseconds. During a quarter cycle (A to B in the sketch), the electrons are accelerated due to the existence of the time-changing flux. The electrons circle the changing magnetic flux many thousand times during each quarter cycle, **acquiring an increment of energy of several hundred electron volts on each revolution.** During this process, incidentally, the linear speed attained by the electrons is very nearly that of the velocity of light.

Near the end of each quarter-cycle, by means of a suitable auxiliary circuit the swiftly moving electrons are caused to spiral away from their orbit and strike a tungsten target, thereby giving rise to intense X rays, which emerge from the tube in a narrow beam.

In Kerst's first machine, which was a relatively small assembly, electrons having energies in excess of 2 mev were made available, with a corresponding X-ray yield. Later, at the General Electric Research Laboratory, Kerst cooperated in the development of a unit which yielded 20-mev electrons; and more recently a much larger betatron has been built, also at the General Electric plant, which is capable of producing 100-mev electrons. Thus it becomes possible to produce, experimentally,

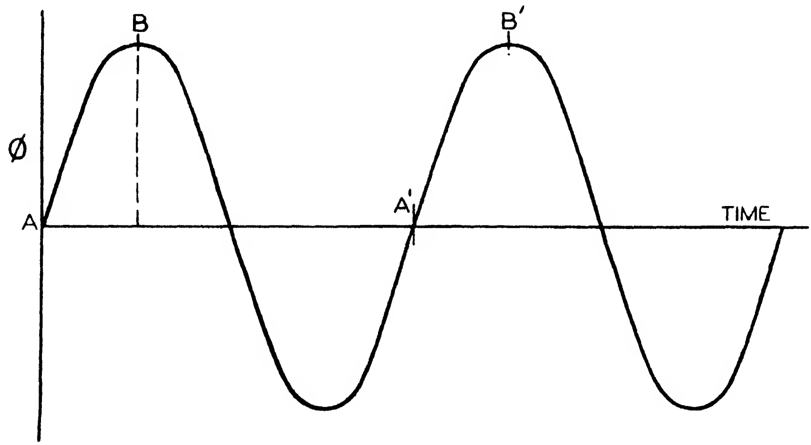


FIG. 330.—Showing the timing of the accelerating pulse as applied in the betatron.

moving electric entities which have a kinetic energy content comparable to that which obtains in the case of cosmic rays. The X rays resulting from the impact of such high energy particles on a target possess penetrating possibilities greatly in excess of any radiation hitherto known. This indicates that we are entering upon an entirely new field of atomic research.

In the latest model of the betatron, the magnet is energized by a 60-cycle current, the magnetic flux density at the orbit having a value of 4,000 gausses. The electrons are injected by applying a potential of 70 kv. The circular electronic orbit is 1.676 m (66 in.) in diameter, and the electrons remain in the changing magnetic field during a quarter cycle. During that time they move about the flux area 250,000 times, and during each revolution acquire an energy increment of about 400 electron volts. The complete unit weighs approximately 130 tons.

The phase-correcting capacitor is the most expensive part of the equipment and probably constitutes the largest capacitor bank now in existence. Brief descriptions of the 100-mev machine are to be found in the *Journal of Applied Physics* for October, 1945, and in the *General Electric Review* for January, 1946. An informative paper on the betatron, by T. J. Wang of Ohio University, appeared in the June, 1945, issue of *Electronics*. In addition to a bibliography of the subject, this paper contains an analytical treatment of the principles involved in the design and operation of this new and highly important scientific tool.

CHAPTER XXVIII

RADIOACTIVITY AND ATOMIC STRUCTURE

218. Discovery of Radioactivity. The discovery of radioactivity in 1896 had its genesis in the discovery of X rays. Since the cathode rays caused various substances to fluoresce and, since these rays had been found to give rise to X rays, the question naturally arose as to whether fluorescence was not in some way connected with the production of X rays. Becquerel,[1] in studying this question, by great good fortune, examined among other substances a number of the compounds of the element uranium. His method was a photographic one and consisted of first exposing a uranium compound to light and then placing it, for a period of several days, near the film of a photographic plate which had been wrapped in lightproof paper. A metal coin was placed between the fluorescent substance and the photographic plate. Upon developing the plate, it was evident that some form of radiation had emanated from the uranium compound and formed a shadow picture on the plate. Further investigation showed that the preliminary exposure to radiant energy had no bearing on the results. Figure 331 is a reproduction of a skiagraph made by Becquerel's method.

In addition to the photographic effects, Becquerel found that the radiation from uranium causes fluorescence, and also gives rise to ionization. Checking his results in various ways he found that the uranium, per se, was the essential substance in the production of the observed effects, and, what is even more important, he learned that the previous condition or treatment of the uranium had no bearing on the result. Fluorescence, for instance, was not a factor in the case. Though on the wrong track, so far as the discovery of a possible source of X rays was concerned, he had, however, been led to uncover a tremendously important truth. Becquerel had, in fact, made a discovery which marked the beginning of an epoch in the advancement of scientific thought; a discovery the results of which were destined to modify profoundly certain fundamental concepts in the domain of both physics and chemistry. Indeed, recent events show that Becquerel's work has been a factor in changing the course of history.

[1] Henry Becquerel, of the Conservatoire des Arts et Métiers, was the son of Edmond Becquerel and the grandson of A. C. Becquerel, both of whom were eminent physicists.

As soon as Becquerel announced his findings, other investigators promptly and vigorously took up the search for other bodies which might possibly possess the recently discovered property exhibited by uranium, a property which soon came to be designated by the term **radioactivity**.

In 1898, Schmidt discovered that the element thorium is radioactive, and to about the same degree as uranium.

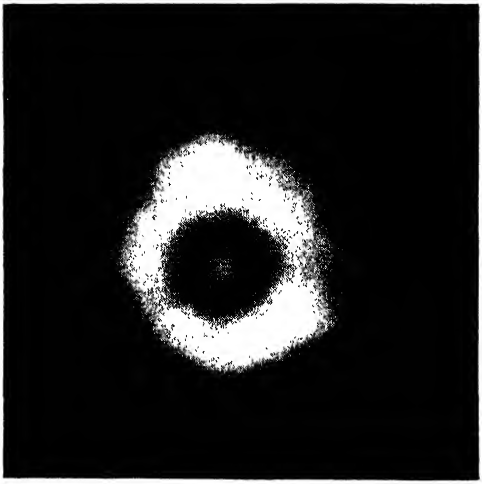


Fig. 331.—Autoskiagraph made by a uranium ore (carnotite). Note the shadow cast by a copper washer.

219. Discovery of Radium. Among the workers who were studying in the new field we find M. and Mme. Curie, who had begun a series of laborious but brilliant experimental researches. The Curies proved that the radioactivity of uranium salts is directly proportional to the uranium content, thus establishing the fact that the radioactivity of uranium is an atomic property.

As a result of an examination of the natural minerals from which uranium is derived, the Curies found that certain specimens of pitchblende were more radioactive than could be accounted for on the basis of their uranium content. It was therefore natural to conclude that the ore contained some other substance which was strongly radioactive. The Curies accordingly subjected a quantity of radioactive pitchblende to a systematic chemical analysis in an attempt to isolate this unknown radioactive body. Barium is present in pitchblende, and it was found on precipitating that element as the carbonate that the precipitate was strongly radioactive. It was, however, found to be impossible to make any further separation by means of reagents, so recourse was had to the method of fractional crystallization. The barium was converted to the

chloride and, as the fractional crystallization progressed, those parts which separated out first proved to be more and more radioactive. By a long and extremely laborious procedure, M. and Mme. Curie were thus able to secure a concentrated radioactive product whose activity was of the order of a million times the activity of uranium. The end product of the analysis was a chemically pure body, which proved to be a new element which they elected to call **radium.** It is interesting to note, in passing, that it was necessary to treat chemically several hundred kilograms of pitchblende ore in order to secure a few milligrams of the element radium. Radium was found to be an element having chemical properties closely resembling those of barium, calcium, and strontium, and having an atomic weight of 225.97. The element has been reduced to the solid form, though it is commonly handled in the form of the bromide (RBr_2) or the chloride (RCl_2). As a metal it is silver white in color, tarnishes rapidly in air, and melts at 700°C. Its radioactive properties will be discussed later.

While working on the preparation of radium from pitchblende, Mme. Curie was successful in separating from the antimony-bismuth chemical group a second radioactive body, which is several times more active than uranium and which she named **polonium,** after her native country, Poland. In more recent years polonium has been found to be identical with Ra F, which later will be shown to be one of the disintegration products of radium. Debierne and Giesel also succeeded in obtaining from the iron group in pitchblende still another radioactive substance, known as **actinium.** The radioactivity of actinium is comparable to that of radium. In more recent years several other elements have been found which exhibit natural radioactivity, but only to a minor degree.

220. The Nature of Radioactivity. In general, there are three types of "rays" emitted from a radioactive body; they were designated by Rutherford as α-, β-, and γ-rays. Let us now consider the nature of each type of emission, seriatim.

If a quantity of strongly radioactive substance is placed in a lead block, as shown in Fig. 332, and a strong magnetic field is brought to bear on the emission, the direction of the field being at right angles to the plane of the figure, the rays will be separated into three parts as shown, and the several groups of rays can be separately detected and identified. A part of the emission is strongly deflected, and the direction of deflection indicates that the entities which go to make up this part of the beam of radiation manifest a negative charge. These are the so-called beta (β) rays.

Another part of the radiation is deflected to a lesser degree by the magnetic field, and the direction of movement shows that the bodies

composing this part of the radiation carry a positive charge. They are the alpha (α) rays.

A third portion is undeflected and shows no evidence of an electric charge. These are the gamma (γ) rays.

Without going into detail concerning the method whereby the data have been secured, the properties and nature of the three types of rays may be summarized as follows.

The **α-particles** are now known to be helium nuclei, the original atom having lost its two planetary electrons. (See nuclear chart, page 3.) The velocity of these entities (positive ions) depends upon the particular radioactive material from which they are ejected, the values ranging from about 1.42×10^9 to 2.05×10^9 cm/sec. The corresponding energies are 4.19 and 8.78 mev, respectively. Though the energy content is seen to be large, the α-particles show the least penetrating power of the three types of rays. Indeed, they would be stopped by the sheet of paper on which these words are printed, or by an aluminum sheet 0.06 mm in thickness. Their "range" in air, under standard conditions, is about 7.5 cm. The α-rays produce marked ionization, being the most effective in this respect of the three types of emission. They also give rise to fluorescence, particularly in the diamond and in zinc sulphate. The α-particle shows a positive charge whose magnitude is **twice** that of the charge carried by the hydrogen ion in electrolysis. The kinetic energy of a moving α-particle is sufficient to disrupt an occasional atom of certain elements. Because of this fact, this entity has been utilized as a "projectile" for the purpose of detaching electrons from atoms, and thus bringing about transmutation.

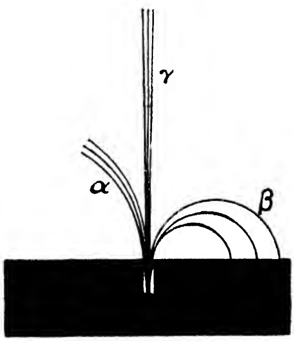


FIG. 332.—Behavior of the emission from a radioactive substance when subjected to a magnetic field whose direction is normal to the plane of the figure.

β-rays are identical with the cathode rays in a vacuum tube; they consist of electrons. However, the velocity of β-rays is greater than that of cathode rays, being of the order of 10^{10} to 3×10^{10} cm/sec. As they are spontaneously ejected from a disintegrating atom, the β-entities show energy values ranging from zero to as much as 10 mev. Most β-rays, however, do not have energy values above 2 or 3 mev. It will thus be seen that the β-rays, in general, have less energy than the α-rays. β-rays produce ionization but not to the same extent as α-particles. Since less energy is spent in the ionization process, β-rays have a much greater range than α-particles, the range in air being several meters.

Their penetrating power is also great; they will pass through as much as a millimeter of aluminum. β-rays produce fluorescence when incident on platino-barium cycanide and certain other substances.

γ-rays constitute the only part of the emission from radioactive bodies to which the term ray strictly applies. It has been established that the gamma radiation is essentially the same as X rays; in general, however, they are more penetrating than X rays. The hardest X rays are completely stopped by 3.2 mm of lead, but γ-rays will pass through 30 cm of the metal. This part of the emission from radioactive bodies is, in short, a periodic electromagnetic disturbance in the ether, and, except in the matter of frequency, is the same as visible radiation. The wave length of γ-rays depends upon the nature of the radioactive material from which they originate. The wave-length values range from 3.9 to 0.0466 Å. The corresponding photon energies would be 0.032 and 2.65 mev, respectively. Thus it is seen that certain of the γ-rays are shorter than the X rays commonly produced. However, by means of the betatron (Sec. 217), X rays may now be produced which are comparable in frequency to the hardest γ-rays. It is important to note that **γ- and β-rays always occur together,** while α-rays may exist alone. Gamma rays produce marked ionization and also affect a photographic plate, as would be expected from their nature.

221. Methods of Measuring Radioactivity. Before proceeding further it will be well to glance at the means employed in the estimation of the degree of radioactivity which a given substance exhibits.

It has already been noted that radioactive bodies produce photographic effects, give rise to fluorescence, and bring about ionization. The last-mentioned effect is the one most commonly utilized in making quantitative measurements of radioactivity. M. and Mme. Curie employed the ionizing property of the substance under test as a measure of its radioactivity, and made use of the gold-leaf electroscope as an indicator of the degree of ionization. Figure 333 is a diagrammatic sketch of the apparatus used by the Curies in their original investigation. If the radioactive material be spread on the plate P, the electroscope first having been charged to a given potential, the air between the plates P and P' will be ionized and the electroscope gradually discharged; the rate of discharge will be a function of the degree of radioactivity. The rate of movement of the gold leaf is observed by means of a suitable optical arrangement. Various other forms of gold-leaf electroscopes have been devised for use in connection with radioactive investigations, one designed by Wilson and improved by Kaye being particularly sensitive and convenient.

The quadrant and other forms of electrometers are made use of in

ionization investigations. In more recent times some form of the Geiger-Muller counter (Sec. 208) has come to be widely used in connection with radioactive studies.

As we shall see in the next section, radioactivity involves atomic disintegration, and the process by which this breaking down of the atomic structure takes place, as well as the resulting phenomena, have long been the objects of intensive investigations. A piece of apparatus which is extensively used in the study of ionization and related phenomena is known as the **Wilson cloud chamber.** Shortly after the discovery of radioactivity, C. T. R. Wilson, an English physicist, developed a simple but effective method of making the results of the ionization process directly visible.

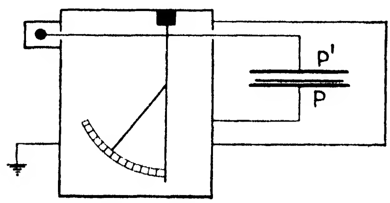


FIG. 333.—Sketch of the electroscope used by the Curies in their original investigations.

It has long been known that small material bodies, such as particles of dust, will serve as nuclei about which water vapor will condense from an atmosphere which is at or near the dew point. Professor Wilson found that gaseous ions may also be made to serve as condensation nuclei. In its simplest form the Wilson condensation apparatus consists of a chamber (Fig. 334) having a tight-fitting piston. A trace of water in the chamber produces a saturated atmosphere at ordinary temperatures. If the piston P is lowered very rapidly adiabatic expansion of the gas is brought about, with a consequent cooling of the gas. As a result, a state of supersaturation comes to exist, and the water vapor will tend to condense on any ions which may be present, thus making the position of the ions visible. If R be a source of α- or β-particles these entities will produce ionization as they are emitted from the radioactive body. As the α-particles, for instance, shoot through the gas in the chamber they will ionize some of the gas atoms that they strike in their flight. The resulting ions will then serve as condensation nuclei, and the trajectory of the ionizing entity is thus made clearly visible. Indeed, the whole ionizing process may be studied visually or photographically by means of this piece of apparatus. If a magnetic field is brought to bear on the region within the chamber the paths of the ionizing entities

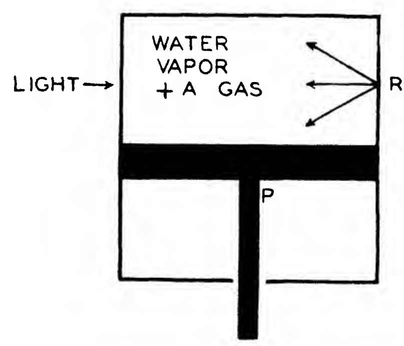


FIG. 334.—Wilson cloud chamber.

will be curved, and from these curves their velocity and energy can be computed.

In the more recent models of this equipment the piston is replaced by a flexible diaphragm, and alcohol vapor is used instead of water vapor. Provision is also made whereby ionizing rays of various types may be shot through the wall of the chamber. Gases other than air are also utilized. In some installations conditions are so arranged that the entrance of the ionizing particle into the expansion chamber automatically brings about the expansion process, and also makes the photographic exposure.[1] It is not an overstatement to say that the Wilson cloud chamber has proved to be one of the great research tools in the field of modern physics. Figure 335 shows typical ionization records made by the condensation method just described. In a later chapter reference will again be made to the Wilson device.

FIG. 335.—Alpha-ray tracks in a Wilson cloud chamber. An α-particle has collided with an oxygen atom. Note also the deflection due to a near hit. (*From "Radiation from Radioactive Substances" by Rutherford, Chadwick, and Ellis, Cambridge University Press, England. By permission of The Macmillan Company, publishers, N. Y.*)

222. Radioactive Transformations and Their Significance. That changes are taking place in radioactive bodies, particularly in the case of radium, is evidenced by the fact that radium compounds are always at a higher temperature than their surroundings, thus indicating that thermal energy is constantly being spontaneously liberated. A portion of radium compound may be as much as 2°C above the temperature of its surroundings. It has been estimated that 1 gm of radium liberates heat at the rate of approximately 130 gm-cal/hr. This is sufficient heat to raise 130 gm of water 1°C and, so far as observations on samples of radium have gone, this liberation of heat has shown no appreciable diminution. This appears to be an anomalous phenomenon, but is found to be explicable in the light of results secured in recent investigations, as we shall see shortly.

Basing their reasoning on a large amount of experimental data, Rutherford and Soddy, in 1902, brought forward a theory which is to the

[1] For descriptions of several modern forms of the Wilson cloud chamber see articles by J. C. Street and E. C. Stevenson, *Rev. Sci. Instruments*, **7**, 347 (1936); G. L. Locker, *Rev. Sci. Instruments*, **7**, 471 (1936).

effect that the radioactive elements are in a more or less unstable condition, the atoms undergoing a process of disintegration. This disintegration or transformation process follows a perfectly definite law. It has been found that radioactive elements of higher atomic weight, as the result of the spontaneous emission of α-, β-, and γ-rays, change to other radioactive elements of lower atomic weight. There are three distinct series of such changes, these groups being commonly referred to as the uranium series, the actinium series, and the thorium series. The uranium series involves radium and may be taken as typical. The following table serves to give a graphic representation of the transformation steps, beginning with uranium and ending with one of the isotopes of lead.

From an examination of the chart it will be seen that uranium is the original parent body of radium and that a form of lead known as radium-lead is the stable end product of the complete series. Whenever an

URANIUM SERIES

Substance	Particle emitted	Atomic number	Atomic weight	Half life
Uranium I	α	92	238	4.56×10^9y
Uranium X_1	β	90	234	24.1d
Uranium X_2	$\beta\gamma$	91	234	1.14m
Uranium II	α	92	234	2.7×10^6y
Ionium	α	90	230	8.3×10^4y
Radium	$\alpha\beta\gamma$	88	226	1590y
Radon	α	86	222	3.825d
Radium A	α	84	218	3.05m
Radium B	$\beta\gamma$	82	214	26.8m
Radium C	$\alpha\beta\gamma$	83	214	19.7m
Radium C′	α	84	214	10^{-6}s
Radium C″	β	81	210	1.32m
Radium D	$\beta\gamma$	82	210	22y
Radium E	$\beta\gamma$	83	210	5.0d
Radium F	α	84	210	140d
Radium G	....	82	206	Stable lead

α-particle (positive helium ion) is expelled from a given element, the atomic weight of the transformation product is less by four than its parent body, four being the atomic weight of helium. Radium itself comes into being when an ionium atom expels a single α-entity, that is, a helium ion. In common with its atomic "ancestors," radium itself is an unstable element and, as a result of its atomic disintegration, a helium

ion (α-ray) is expelled. The remainder of the atomic "wreck" constitutes an atom of gas, sometimes spoken of as **radon,** and often referred to as **radium emanation.** The element radon is a relatively heavy inert gas that boils at 65°C; it may be liquified at -150°C. A gaseous element corresponding to radon is also evolved as one of the steps in the transformation which occurs in both the actinium and thorium series.

In the actinium series the end product is also an isotope of lead, whose atomic weight is 207. The thorium family yields another isotope of lead; atomic weight 208. We have cited the particular case of the radioactive element radium and its emanation, radon, and the product helium, in order to call attention to what may be considered a transformation, or, as it would have been called in the days of the alchemists, the transmutation of certain elements. Perhaps the most remarkable thing about this strange transformation process is that it is entirely beyond our control; it proceeds quite independently of local physical and chemical conditions. In such a process of atomic transformation we have a change that transcends hitherto known chemical laws.

It should be noted that these radioactive changes follow perfectly definite laws. The rate of disintegration of any radioactive element may be expressed by the relation

$$M_t = M_0 \epsilon^{-\lambda t},$$

where M_0 represents the initial mass of radioactive material; M_t the mass existing t seconds later; ϵ the base of the Napierian logarithms; and λ the transformation constant for that particular material, *i.e.*, the proportion of active material which undergoes change each second. Since the above relation is an exponential expression, it has been found convenient to deal with what is known as the **half period of transformation;** that is, the time required for one-half of the particular body to be transformed. If we modify the above expression to make it cover this half-period concept it becomes

$$\frac{M_0}{2} = M_0 \epsilon^{-\lambda T},$$

where T is the time required for one half of the substance to be transformed. This reduces to

$$T = \frac{\log \epsilon^2}{\lambda} = \frac{0.6931}{\lambda}. \tag{269}$$

The values of the half-period T are indicated for each radioactive body in the uranium series shown in the preceding table. The half-period for radium has been determined to be about 1,600 yr, while for radon (radium emanation) it is only 3.83 days. To put the case differently, a quantity

of radium will not depreciate more than 4 per cent in 100 yr. The fact that the disintegration process is so slow in the case of radium explains why the evolution of heat by a radium salt has not shown any appreciable diminution during the period of time over which observations have thus far been made.

223. Artificial Transformation of Elements. While only a comparatively small number of the elements are naturally radioactive, it has been found possible to decompose **artificially** many of the nonradioactive elements. The method by which this is accomplished is both interesting and significant.

It was observed by Rutherford that when a group of α-particles is caused to pass through a gas or a thin metal foil a few of them are sharply deflected at some point in their path. Such deviations from an otherwise straight path are shown in records of α-ray tracks in Fig. 335. From his observations Rutherford concluded that most of the α-projectiles pass directly through an atomic structure without encountering any of the constituent parts of the atom. However, an occasional α-particle apparently passes near enough to a nucleus to be affected by the repelling electrostatic force due to the nuclear charge; hence the abrupt change in its path. Further, there was evidence that an α-particle does occasionally actually strike a nucleus, with the result that a part of the nucleus is detached and moves out of the atom with a speed comparable to that of the impinging α-particles. The representative record, reproduced as Fig. 336, gives evidence that such an event takes place.

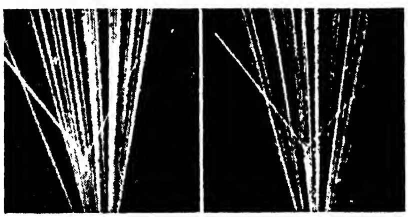


FIG. 336.—Collision of an α-particle with a helium atom. (*From "Radiation from Radioactive Substances" by Rutherford, Chadwick, and Ellis, Cambridge University Press, England. By permission of The Macmillan Company, publishers, N. Y.*)

In 1919 Rutherford carried out an investigation the results of which clearly demonstrated that the artificial disintegration of matter can be brought about by bombarding the atoms of an element with α-particles. He first worked with hydrogen, nitrogen, and oxygen. In the case of nitrogen the detached part of the nucleus was found to be a high-speed **H nucleus,** *i.e.*, a **proton.**

A few years after the above-mentioned discovery, further studies along the same line were made by Rutherford and Chadwick. They secured evidence that all elements from boron to potassium, except carbon and oxygen, can be disintegrated by high-speed α-particles, and in all cases **a high-speed H nucleus, or proton, made its appearance.**

Thus it was established that the hydrogen nucleus, or proton, is one of the fundamental units that enter into the composition of an atomic nucleus.

It is interesting to note, in passing, that in 1815, Dr. Prout, an English physician, suggested that hydrogen was the basic material from which all elements are made. The facts outlined above indicate that Prout's hypothesis contained an important element of truth, though his original suggestion was rejected at the time.

The experimental findings of Rutherford and Chadwick were of the utmost importance; the way was opened for an extended and fertile series of researches into the constitution of matter. It may indeed be said that these experiments at Cambridge mark the beginning of what has now come to be designated as **nuclear physics.**

224. The Neutron and the Positron. In dealing with the subject of nuclear physics one must note two important discoveries which occurred in the year 1932. Prior to that date it was generally accepted that an atomic nucleus consisted solely of protons and electrons. But during that eventful year certain facts came to light which resulted in a revision of that explanation.

In 1931, two German investigators, Bothe and Becker, found that the bombardment of certain of the lighter elements, such as beryllium, lithium, and boron, with α-particles gave rise to a very penetrating type of secondary emission. Indeed, it was found that this emission was more penetrating than the hardest γ-rays. It was at first assumed that in some way the alpha bombardment had resulted in the emission of an extremely short-wave type of gamma radiation. However, the energy relations involved in the process appeared to point to a different explanation of this new phenomenon. In France the Curie-Joliots (1932) observed that this emission was capable of causing the ejection of protons from paraffin and other materials containing hydrogen. Chadwick carried out a series of experiments at the Cavendish laboratory which led him to suggest that the results secured by previous investigators could be explained on the basis that the newly observed emission from beryllium and other bodies, when bombarded with α-particles, consists of particles whose mass is approximately equal to that of the proton but which **carry no charge.** This viewpoint has now been thoroughly established, and these new entities have been named **neutrons.** The reaction by which the neutron comes into being, in the case of beryllium, is

$$_4Be^9 + {_2He^4} \rightarrow {_6C^{12}} + {_0n^1},$$

where $_0n^1$ represents the neutron, having an atomic weight of 1 and an atomic number zero, since its charge is zero. Here we have a case where the impinging α-particle (He nucleus) **is captured,** and becomes the

nucleus of a carbon atom. Other reactions of a similar nature are now known to yield neutrons.

Neutrons possess very high energies, some from beryllium showing a value as high as 13.7 mev. Because of their high energy content, and because they are electrically neutral, neutrons have proved to be very useful as projectiles with which to bombard other nuclei. In passing, it may be added that Dr. Chadwick was awarded the Nobel prize in 1935 for his identification of the neutron, and for other important related research.

In a previous discussion it was noted that the phenomenon of radioactivity was discovered as a result of research having to do with X rays. Another discovery has a somewhat similar history. Later (Sec. 228) we shall consider the subject of cosmic rays. For the present it may be said that some form of highly penetrating "rays" come to the earth from outer space. It had been observed that these extraterrestrial rays, when absorbed by matter, caused the ejection of high-energy charged entities from the absorbing atoms. At the research laboratories at the California Institute of Technology it was felt that one might secure some information about the nature of cosmic rays if a study was made of the secondary emission resulting therefrom. Accordingly, a special Wilson cloud-expansion chamber was constructed, in the assembly of which was incorporated an electromagnet capable of developing a field of some 20,000 gausses. So strong was this field that it was capable of producing an appreciable deflection in the trajectory of particles having energies of the order of 500 mev. On several of the photographic records made with this equipment by Dr. C. D. Anderson ray tracks appeared which, because of the direction of their curvature, indicated that they were made by **positively** charged particles having a mass equal to that of the electron, with which we have long been familiar. The photograph which gave the most conclusive evidence that such an entity actually appeared is reproduced as Fig. 337. A lead partition 6 mm thick was placed across the expansion chamber. It is to be noted that a particle entering at the upper edge passes entirely through the plate, the curvature of its path being much greater below the partition than above. This means that it lost energy as it traversed the dividing wall. Measurements show that its energy was 63 mev above the plate and 23 mev below. From the direction of curvature, and from the energy relations just given, Dr. Anderson concluded that he was dealing with a hitherto unknown entity, **the positive counterpart of the negative electron.** The new particle has come to be called the **positron,** its existence having been confirmed by various other workers at home and abroad. The photograph shown as Fig. 337 is of great historical importance. Dr. Anderson's

discovery constitutes one of the great advances in modern physics, and for it he was awarded the Nobel prize. And thus the study of cosmic rays led to the discovery of a new physical entity, adding one more particle to the list with which the physicist must deal.

Various questions naturally arise in connection with the study of positrons. Are they numerous? Have they an independent existence? Do they form a part of atomic nuclei?

It must suffice, here, to summarize briefly the answers to these questions. It seems probable that positrons readily combine with electrons, and hence their independent existence is extremely short—a small fraction of a second. There is evidence tending to show that electron positron pairs may appear at the point where a γ-ray photon impinges on a nucleus.

FIG. 337.—Cloud chamber record of the trajectory of a positron, as made by C. D. Anderson in 1932. As viewed in the illustration, the positron was moving downward. Note that after passing through a 6-mm lead plate its curvature is changed, indicating that the particle has lost a definite amount of energy. Knowing this loss of energy and the direction of the magnetic field it becomes evident that the particle is a positive entity. This is a photograph of great historical interest.

One is next moved to inquire as to what happens when an electron and a positron combine. Existing evidence tends to indicate that **when such a union occurs the two masses disappear and simultaneously a definite amount of radiant energy makes its appearance.** It will be obvious that the foregoing statements are revolutionary in nature—energy being transformed into matter and matter being transformed into energy. And yet, on the evidence at hand, we are forced to such a conclusion.

225. Artificially Induced Radioactivity. Scientific events sometimes move forward at a rapid pace. It will be recalled that the positron was identified by Anderson in 1932. It was soon found that positrons could be had by bombarding some of the lighter elements with α-particles. In 1933 the Joliots were experimenting with the positrons which are obtained by the bombardment of boron, magnesium, and aluminum with α-particles from polonium. These investigators observed that **the bombarded substance continued to emit positrons after the source of α-particles had been removed.** Further, this post-excitation emission

was found to decrease with time, the decay in activity following an exponential law. It is thus evident that we have here not only a process of transmutation, but also **the production of artificially induced radioactivity.** In January, 1934, M. and Mme. Curie-Joliot announced the discovery of **induced radioactivity,** and the scientific world moved one step farther along the road toward a better understanding of the nature of matter. It is interesting to note that the daughter of one who spent her life in the study of radioactivity should be one of the co-discoverers of artificial radioactivity. For this highly important discovery, the Joliots, in 1935, were jointly awarded the Nobel prize in chemistry for that year.

The reactions that take place in the production of induced radioactivity usually consist of two steps. The case of aluminum is typical, thus,

$$_{13}\mathrm{Al}^{27} + {}_{2}\mathrm{He}^{4} \rightarrow {}_{15}\mathrm{P}^{30} + {}_{0}n^{1}$$

and

$$_{15}\mathrm{P}^{30} \rightarrow {}_{14}\mathrm{Si}^{30} + {}_{1}e^{0}.$$

It will be seen that the first stage of the process involves the capture of a helium nucleus ($_{2}\mathrm{He}^{4}$) and the formation of an unstable isotope of phosphorus plus the liberation of a neutron. In the second stage, the phosphorus is transformed into a stable isotope of silicon with the emission of a positron ($_{1}e^{0}$). The half life of the radioactivity in this particular case is 2.5 minutes. It has been found that induced radioactivity is frequently accompanied by the ejection of electrons instead of positrons. It is now known that radioactivity may be induced by bombarding the nuclei of atoms by high-speed entities other than α-particles, as we shall see shortly. As the result of the intensive and extensive investigations carried on by Fermi and others, some 300 artifically radioactive isotopes have been identified.

226. Production of High-energy Ions for Nuclear Experiments. In the disintegration processes thus far described some one of the emission products of a radioactive body was utilized for the purpose of initiating the transmutation. As the study of nuclear physics progressed, the need was felt for projectiles of still higher energy content. Accordingly, the Van de Graaff high-voltage generator was designed. With this machine it has been possible to develop 4.5 million volts. A potential difference of that magnitude makes it possible to impart high energy to electrical entities. The betatron, developed by Kerst and described in Sec. 217, will, in its latest form, yield a supply of 100-mev electrons.

Still another type of installation, originally developed by Lawrence and Livingston at the University of California, is capable of yielding a

copious supply of high-energy ions of several types. Reference is made to the **cyclotron.**

The principle on which the cyclotron operates is comparatively simple. By giving the desired ions repeated impulses, an extremely high speed is imparted to the electrical entities. The method by which the motion of the ions is accelerated at definite space intervals involves the synchronizing of the motion of the ions with an alternating electric field.

The cyclotron assembly consists of a flat hollow cylinder divided into two sections D and D' called the "dees." These cylindrical sections, which function as electrodes, are positioned between the pole pieces of a large electromagnet (Fig. 338). The dees in turn are inclosed in a gas-

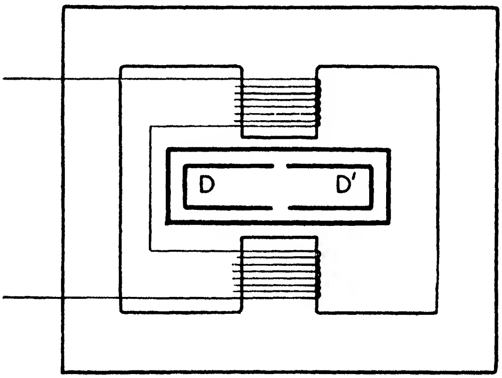


FIG. 338.—General plan of the cyclotron.

tight metal housing. A cathode F (Fig. 339) supplies thermally liberated electrons which are given an initial acceleration by a potential difference that is maintained between the cathode and the metal housing. A h-f alternating potential, of the order of 10^7 cycles/sec, and 10,000 to 20,000 volts, is maintained between the two dees. A gas, say hydrogen, at low pressure is introduced into the chamber. The electrons emitted by the heated filament ionize this gas, and the gaseous ions (in this case protons) will be acted upon by the electric and the magnetic fields.

Let us assume that a positive ion is moving at a moderate speed toward the segment D and that D is at that instant negatively charged. Our ion will then be accelerated toward D. Once this ion enters the region **within** the dee it will not be accelerated, but will move at constant speed. Due to the effect of the magnetic field, its path will be semicircular, and it will eventually reach the space between the dees. If the frequency of the a-c potential is so adjusted that the transit time through the dee is exactly equal to one-half a potential cycle, the ion will reach the edge of the dee just as the potential reverses, *i.e.*, as D' becomes

negative. Acceleration will accordingly take place until the ion enters the second dee D', when its speed again becomes constant. This process continues, the ion acquiring an increment of speed each time it moves across the gap between the dees. Since its speed is increased, the ion will describe a path whose radius is also becoming greater. But fortunately the transit time is constant; hence the ion does not change phase with respect to the a-c potential. The reason why this is so may readily be

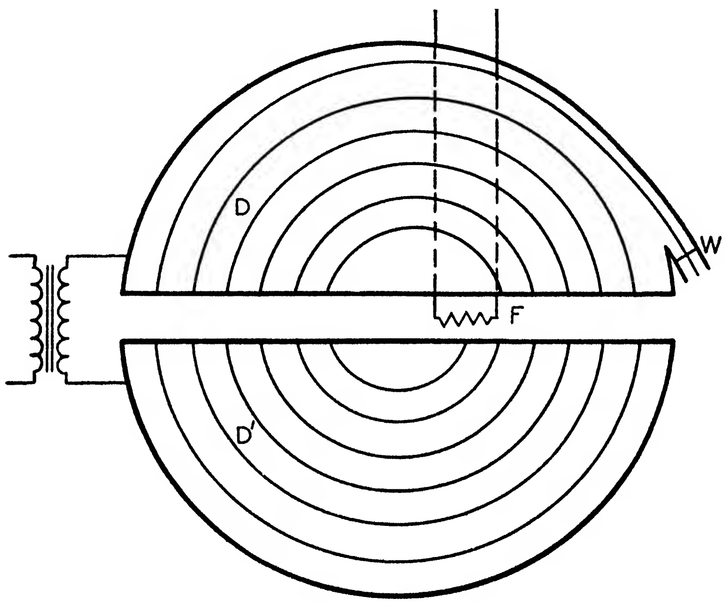


FIG. 339.—Path taken by the ions as they are accelerated within the cyclotron.

seen from the following considerations, based on previous discussions. In Sec. 195 it was shown that when an electron having constant speed is moving at right angles to the direction of a uniform field the radius of its path will be given by the expression

$$r = \frac{mv}{He}. \tag{i}$$

For ions, in general, the charge involved will be some multiple of e; hence (i) would take the form

$$r = \frac{mv}{Hq}, \tag{ii}$$

where the terms have their usual significance. The time required for the

ion to describe a **semicircle** will be

$$t = \frac{\pi r}{v}. \tag{iii}$$

Combining equations (ii) and (iii) we get

$$t = \frac{\pi m}{Hq}, \tag{iv}$$

which shows that, if we neglect any change in mass due to the velocity factor, **the transit time is constant.** Therefore, by adjusting the field strength H, one can cause the ion to arrive at the gap just as the potential changes sign. The ion accordingly will have its speed augmented each time it crosses the gap, and its path will gradually spiral outward, until it finally reaches the outer edge of the dee. A "window" W is provided at that point through which the stream of ions passes into the region outside the chamber. After making a number of complete excursions the ions will have attained a high energy value, and this notwithstanding the fact that the actual accelerating potential is relatively small. An expression giving the kinetic energy of the ions as they emerge from the cyclotron may be readily deduced.

From (iv), above, it follows that the time required for one **complete** trip through the dees will be given by

$$T = \frac{2\pi m}{Hq}. \tag{i}$$

The number of transits per second would then be

$$\nu = \frac{Hq}{2\pi m}, \tag{ii}$$

where ν is the frequency of the alternating potential applied to the dees. The velocity of any given ion as it emerges from the cyclotron would be given by the expression

$$v = 2\pi R\nu, \tag{iii}$$

where R is the final radius. Combining (ii) and (iii) we get

$$v = \frac{RHq}{m}. \tag{iv}$$

From mechanics we have that

$$\text{Kinetic energy} = \tfrac{1}{2}mv^2. \tag{v}$$

Substituting the value for the velocity given by (iv) in (v) there results

$$\text{K.E.} = \frac{R^2H^2q^2}{2m} \quad \text{ergs.}$$

To express the above relation in terms of electron volts we may multiply by the proper transformation factor and get

$$\text{K.E.} = \frac{R^2H^2q^2 \times 6.24 \times 10^5}{2m} \quad \text{mev} \tag{270}$$

as the final expression for the energy content of the emerging ions. Equation (270) shows that the energy acquired by the ions as they spiral through the cyclotron is porportional to the square of the radius of the dees. In order to secure high-speed ions it is therefore desirable to have the diameter of these electrodes as large as possible. This involves a correspondingly large electromagnet. One of the largest cyclotrons now in operation has 60-in. pole faces; a still larger unit is under construction.

In using the cyclotron, if a supply of high-speed deuterons is wanted, heavy hydrogen (deuterium) is introduced into the chamber instead of hydrogen. To secure α-particles, helium is used in the ionization chamber. Thus the cyclotron may serve as a source of high-energy ions of various types, which may be utilized for the purpose of disintegrating other atomic nuclei. In practice, the substance to be bombarded by the issuing high-speed ions is placed near the window through which they emerge. It is now possible to produce 25-mev deuterons and 50-mev α-particles. Thus it becomes possible to produce radioactive sodium, for example, with which to carry on experiments in the fields of chemistry and biology. And much work is being currently done along those lines.

In concluding our brief description of the cyclotron it should be mentioned that Dr. Lawrence was granted the Nobel prize as a result of his outstanding work in the field of nuclear physics.

227. Fission. In 1934, Fermi and his coworkers at the University of Rome investigated the possibility of producing an element having an atomic number higher than 92 by bombarding uranium with neutrons. Several radioactive substances were produced by that process, and it was thought for a time that elements 93, 94, 95, and 96 had actually been produced. However, later work along this line by Hahn and Strassman, in Germany in 1939, culminated in the discovery of a remarkable phenomenon—the splitting or **fission** (as the process has come to be called) of the uranium nucleus into **two parts having nearly equal masses** when subjected to bombardment by neutrons. Apparently, upon capturing a neutron, the nucleus of the uranium atom becomes unstable and separates into two fragments, one of which, at least in certain cases, proves to be the radioactive barium isotope $_{56}Ba^{139}$, and the other $_{57}La^{140}$. There is found to be an excess of neutrons after the disintegration occurs. This constitutes a highly important fact. These secondary neutrons are either immediately emitted as such, or their number is diminished as a

result of electron emission. If these excess neutrons possess sufficient energy, we are faced with the possibility of an extremely rapid chain reaction, triggered off by the capture of a single extra-atomic neutron. A further interesting aspect of this unique process at once presents itself. The total mass of the final disintegration products is **less** than that of the original nucleus plus the captured projectile.

It will be worth while to digress for a moment to consider the energy aspects of a nuclear reaction.

If we take the simple case where an α-particle is caused to strike the nucleus of a sodium atom, the situation may be represented thus,

$$Na^{23} + He^4 \rightarrow Mg^{26} + H^1.$$

Substituting the atomic masses we have

$$22.99610 + 4.00389 = 26.99999$$

and

$$25.98980 + 1.00813 = 26.99793.$$

It is thus evident that 0.00206 unit of atomic mass has ceased to exist **as mass.** What is its fate?

In developing his theory of relativity Einstein introduced the concept of the equivalence of mass and energy. According to this point of view a given mass m is equivalent to a definite amount of energy E. The mathematical statement of this principle takes the form

$$E = mc^2,$$

where c is the speed of light. This is one of the revolutionary principles of modern physics. If we apply Einstein's relation to the reaction cited above, on the basis that 1 atomic mass unit is equivalent to 931 mev, it is found that 1.92 mev of energy makes its appearance when 0.00206 unit of mass **ceases to exist as mass.** A corresponding release of energy occurs whenever the total atomic mass of the final product in a nuclear reaction **is less than that of the original particles.** In ordinary nuclear disintegrations quantities of energy varying from 1 to 20 mev are set free. Experimental findings confirm the theoretical calculations, thus verifying the principle enunciated by Einstein.

However, in the case of the fission of the uranium nucleus an extremely large quantity of energy is released. Henderson has found that the energy made available from a single uranium fission is of the order of 200 mev. Flügge has estimated that 1 m^3 of U_3O_8 might develop 10^{12} kilowatt-hours of energy in less than 0.01 sec. All of this energy would probably not be available in kinetic form; some undoubtedly would be absorbed in nuclear processes that accompany the fission. However,

there would still remain enormous quantities of energy which might be put to use if it could be controlled. That such a source of energy can be tapped and utilized is attested by the devastating results attained by the use of the now famous atomic bomb. While little detailed information[1] is available at this time the data which have been released would appear to indicate that the nuclear reaction involved in the atomic bomb disintegration might be written as

$$_{92}U^{235} + {}_0n^1 \rightarrow {}_{92}U^{239} \rightarrow {}_{93}Np^{239} + {}_{-1}e^0$$
$$_{93}Np^{239} \rightarrow {}_{94}Pu^{239} + {}_{-1}e^0,$$

where Np and Pu represent two new (transuranic) elements called neptunium and plutonium, respectively.

Prior to the World War I it was known that three elements might be caused to undergo fission: uranium, thorium, and protoactinium. In the case of uranium, the 235 isotope appears to be the most usable form of that element. This splits into barium and krypton, with the liberation of three neutrons. It is known that fission may be brought about by bombardment by deuterons, α-particles, neutrons, and protons, and by irradiation by γ-rays.

For a comprehensive review of the subject of nuclear fission, the student should consult a paper by L. A. Turner which appeared in the January, 1940, issue of *Review of Modern Physics*.

228. Cosmic Rays. At the beginning of World War II the two most active fields of research, so far as physics was concerned, were nuclear physics and cosmic rays. In the preceding sections of this chapter we have briefly reviewed a few of the more important aspects of nuclear research, and in doing so have referred incidentally to the subject of cosmic rays. Because of the vast amount of research effort expended in cosmic-ray studies, our limitations will permit only a brief outline of the most significant results which have been brought to light.

It was long known that a charged electroscope would gradually lose its charge, even when every precaution was taken to guard against all known causes. In seeking to determine the cause of this phenomenon, Rutherford and McLennan in 1903 observed that the rate of discharge was lessened if the electroscope was inclosed in a heavy metal housing. This appeared to indicate that the ionization causing the discharge was not due to local causes, but rather that it might possibly be caused by some form of radiation from a distance. The subject was pursued further by a German physicist, Gocken, who in 1910 and 1911 took an

[1] Such public information as is now (Sept. 1945) available is contained in a report prepared by Professor Henry D. Smyth of Princeton University entitled "Atomic Energy for Military Purposes."

electroscope aloft in a balloon to an altitude of 13,000 ft. He found that there was little if any diminution in the ionization at that height. During 1912 to 1914 two other German investigators, Hess and Kolhörster, repeated the balloon experiments, going to a height of 5.6 miles (9 km). They found that the ionization decreased slightly for the first two miles and then **increased** decidedly with altitude, reaching a maximum value of eight times what it was at the earth's surface. It thus became evident that the ionization which manifests itself by the discharge of an electroscope is due to some form of radiation reaching the earth from outer space. Hess was the first to formulate definitely such an explanation; and he further pointed out that since this radiation fell upon the earth from all directions, its origin could not be the sun. Since the origin of the rays is extraterrestrial and is not from the sun, they have come to be designated by the term **cosmic rays.** Where they actually originate, or how, is not known.

Once it was established that such a form of radiation existed the next question to receive the attention of investigators had to do with the **character** of the rays. Two groups of investigators have carried on extensive studies into the nature of cosmic rays: one group headed by Dr. Millikan and the other under the direction of Dr. A. H. Compton. Dr. Millikan and his coworkers not only explored the upper atmosphere but also submerged electroscopes in snow-fed lakes. They found that the rays were capable of penetrating 72 m of water, which would be the equivalent of nearly 2 m of lead. Experiments carried out in mines by other investigators indicate that some of the rays are capable of penetrating a layer of earth equivalent to 124 m of lead. It is thus apparent that cosmic rays are much more penetrating than the hardest γ-rays. On an energy basis, cosmic rays would be rated in **billions** of electron volts, rather than in millions, as in the case of γ-rays.

Probably the most comprehensive single investigation ever undertaken prior to World War II was led by Dr. A. H. Compton while at the University of Chicago. Some 80 physicists, at 100 widely separated places, collaborated with Dr. Compton in securing data bearing on cosmic rays. Twelve different expeditions were involved. Later Compton and Turner carried out an extensive series of cosmic-ray measurements with greatly improved equipment. These later studies involved 12 trans-Pacific steamship voyages between Vancouver, Canada, and Sydney, Australia, and occupied the greater part of a year.

The most important finding of these extensive investigations is that cosmic-ray intensity is a function of magnetic latitude. Figure 340 is a reproduction of the curve made from data secured by Compton and Turner, and by Gill, who later made 15 additional trans-Pacific voyages.

The significance of the graph is obvious; the magnetic field of the earth produces an effect on this strange system of rays reaching our planet from outer space. This fact, together with other evidence, points to the

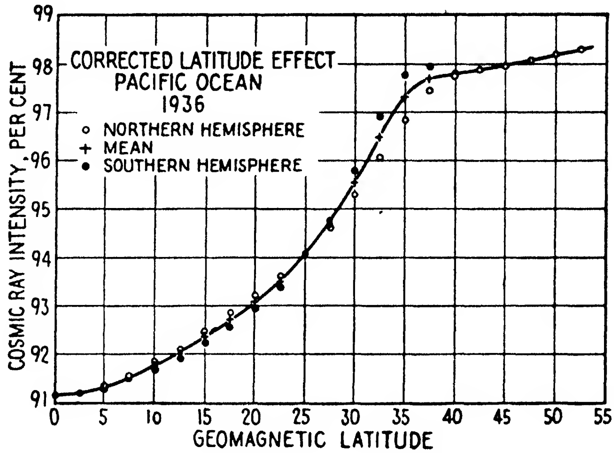


FIG. 340.—Showing the effect of magnetic latitude on cosmic ray intensity. (*Courtesy of Physical Review.*)

tentative conclusion that **cosmic rays are not photons,** as first suggested by Millikan, but that they consist largely of charged entities, such as electrons, positrons, and mesotrons (Sec. 229). Cosmic radiation appears to consist of two parts: a soft part, probably made up of electrons and positrons; and an extremely penetrating component consisting of meso-

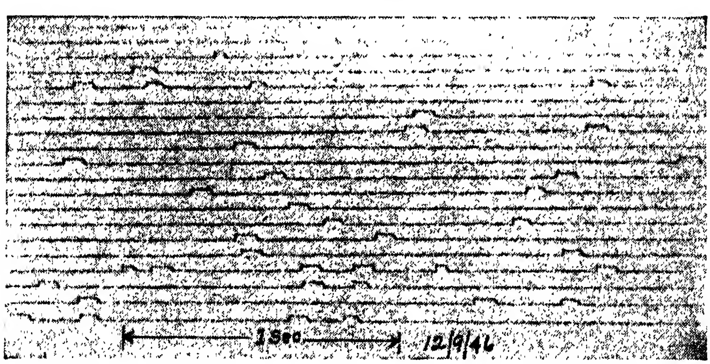


FIG. 341.—Typical record of cosmic-ray pulses as they occur at Northfield, Minn

trons. The arrival of these entities is quite irregular in occurrence, as indicated by the automatically made tracing (Fig. 341) recorded by the author at Northfield, Minnesota.

While we know practically nothing as to the source of this radiation, its study has led to important discoveries. Nature has provided us with the universe for a laboratory. Perhaps some day the study of cosmic radiation may assist in solving some of the problems of astrophysics.

229. The Mesotron. In the last section reference was made to a particle which we have not yet considered. In 1936 Dr. Anderson, the discoverer of the positron, in studying cloud-chamber records due to cosmic rays, observed effects which indicated the possibility of the existence of an electrical entity different from any particle previously identified. Evidence which has accumulated since Anderson's original observations were made clearly indicates that a particle does exist that has a mass something like 200 times the mass of the electron, and that may show either a positive or a negative charge. Thus the new member of the particle family has a mass intermediate between that of an electron and a proton. As previously indicated, the newly discovered particle is commonly referred to as a **mesotron,** sometimes shortened to **meson.**

As yet, our information regarding this entity is rather limited. It is thought by some that mesotrons are produced in the upper strata of the atmosphere by the action of primary cosmic rays. Possibly they ultimately disintegrate in some such manner as do radioactive bodies. Undoubtedly we shall learn more about these entities in the near future.

230. Present Understanding of Atomic Structure. As additional facts have come to light, our earlier understanding of atomic structure has undergone revision, and it is probable that additional research will result in still further modifications of atomic concepts. Currently the picture is about as follows.

The nucleus consists of neutrons and protons, no electrons being normally present in the nucleus. The neutron is electrically neutral; its internal structure is not known. The number of protons (positive entities) present in the nucleus determines the magnitude of the nuclear charge, and hence the atomic number Z. There is evidence to show that some sort of attraction exists between neutrons and protons, but the nature of this force is unknown. In the atom, when neutral, the number of electrons associated with a given nucleus equals the number of protons present. Bohr has thought of these electrons as moving in orbits about the nucleus, and distributed in definite energy levels, sometimes referred to as **shells.** It is coming to be felt that the concept of orbital motion is not necessarily correct, though it is thought that each electron does spin about an axis through itself. We do know that, in general, the outer electrons are the ones most easily detached from the atomic structure. In any event, either the orbital or the shell concept is a convenient way of picturing the situation; such representations have served, and probably

will continue to serve, as useful conventions, in much the same way as Faraday's lines of force. Gradually, however, a mathematical representation of atomic structure and behavior is being evolved, which eventually will probably replace the present conventional scheme of representation.

There is evidence that electron-positron pairs come into existence as a result of the transformation of radiant energy. Possibly the electrons which are ejected from nuclei have their genesis in such a process, **energy disappearing** and **electrons appearing** as a result.

In considering the matter of atomic structure, the question of isotopes presents itself. We have seen that a given element may appear in a single form or in several forms, and that these forms or varieties are designated as isotopes. How are the components of the various isotopes of a given element arranged? Take the case of magnesium, for example. This element has three isotopes, Mg^{24}, Mg^{25}, and Mg^{26}. Each of these bear the atomic number 12, and therefore the nucleus of each contains 12 protons. The Mg^{24} isotope has 12 neutrons in its nucleus; Mg^{25}, 13; and Mg^{26}, 14. Each isotope has 12 electrons, arranged in the same manner in each case. Most of the elements have two or more isotopes, some of which are unstable.

Referring again to the electronic components of the atom, it is apparently true that the electrons may, from time to time, abruptly change their position with respect to the nucleus. In other words, their **energy state** may undergo a change. If a given electron acquires a certain amount of energy from an outside source, it may take up a position farther from the nucleus, *i.e.*, move to a higher energy level. (In terms of the old "orbit" concept, it would jump to an orbit having a larger radius.) This changed condition often appears to be unstable, in which event the electron may, in response to central forces, return to its original position. This results in a liberation of energy in the form of radiation. The concept of this process may be expressed thus,

$$h\nu = w_2 - w_1, \tag{271}$$

where $h\nu$ is the quantum of energy (photon) liberated or absorbed, w_1 the energy content for one state, and w_2 the corresponding energy magnitude when the other condition obtains. On the basis of the above reasoning it is possible to predict the frequency corresponding to the principal spectral lines of some of the more simple elements such as hydrogen.

When we know more about the nature of the neutron, it is probable that we shall be able to arrive at a clearer understanding of atomic structure.

231. Particles versus Waves. In 1900 Planck introduced the quantum theory of radiation. Though none too well received at the time, it is now universally accepted. It will be recalled that Planck's concept is to the effect that radiant energy is emitted in quanta—successive groups of waves, each group possessing a quantity of energy given by $h\nu$. Perhaps the most convincing evidence of the idea that radiant energy in its interaction with matter behaves as if it were composed of discrete entities, is to be found in the Compton effect (Sec. 215). If, then, radiant energy exhibits a dual nature—that of waves and that of particles—might it not be possible that entities such as electrons, protons, neutrons, and even atoms and molecules would show, at least under certain circumstances, the characteristics commonly associated with waves? It was in 1924 that de Broglie, a French physicist, addressed himself to this problem. The hypothesis which de Broglie introduced, revolutionary though it was, has been fully confirmed, and its effects have been far reaching. Beginning with the quantum theory, and making use of certain considerations based on the special theory of relativity, de Broglie derived an expression for the wave length associated with these entities. In its simplest form the development is as follows.

From elementary considerations we know that the quantity of energy assigned to a photon is given by

$$E = h\nu, \tag{i}$$

where h is Planck's constant and ν the frequency of the radiation. The frequency would be given by the expression

$$\nu = \frac{c}{\lambda}, \tag{ii}$$

where c is the velocity of light.

If now we assign to a photon a definite mass m, as may be done on the basis of the relativity theory, the energy of such a mass would be

$$E' = mc^2. \tag{iii}$$

If we equate the expressions for energy given by (i) and (iii) there results

$$mc = \frac{h\nu}{c} = \frac{h}{\lambda}, \tag{iv}$$

where mc is the momentum of a photon whose velocity is c. On the basis of this relation we may express the **momentum** of radiant energy in terms of mass and velocity or in terms of wave length.

De Broglie applied the foregoing reasoning to those entities which are commonly referred to as particles. If the particle has a mass m and

moves with a velocity v, its momentum would be mv. If the particle exhibits wave properties we could equate this momentum to that given by (iv) and obtain

$$mv = \frac{h}{\lambda}.$$

The wave length associated with the particle would be given by

$$\lambda = \frac{h}{mv}. \tag{272}$$

This means that a particle (electron, proton, etc.) having a mass m and a velocity v should have an **equivalent** wave length given by h/mv. The above expression is known as **de Broglie's equation,** and is one of the most important relations in the field of modern physics. Within the past few years de Broglie's original thesis has been expanded by Schrödinger, Heisenberg, Dirac, and others into an elaborate analytical structure known as **wave mechanics**—a subject beyond the scope of this volume. The de Broglie equation, however, is involved in all of the more extended theoretical treatments of the relation between waves and particles.

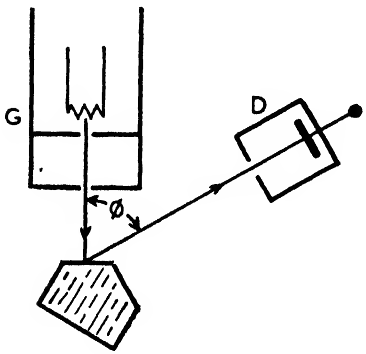


FIG. 342.—Experiment of Davisson and Germer, showing that electrons exhibit wave properties.

The next question that presents itself is this: Has de Broglie's hypothesis been confirmed experimentally? The answer is in the affirmative.

Before describing the first of two confirmatory tests, let us digress for a moment to determine what would be the order of magnitude of wave length that one might expect to find associated with electrons falling through a potential difference of, say, 100 volts. From previous considerations the kinetic energy of such a particle would be given by

$$Ve = \tfrac{1}{2}mv^2.$$

Combining this expression with the de Broglie relation [Eq. (272)] we get

$$\lambda = \frac{h}{\sqrt{2mVe}}. \tag{273}$$

Substitution in this equation gives

$$\lambda = 1.22 \times 10^{-8} \text{ cm} = 1.22 \text{ Å}.$$

It is at once apparent that the wave length is of a magnitude commonly met with in X-ray practice. Since this turns out to be the case, it is obvious that one might check this computation by measuring the electron wave length using a crystal as a diffraction grating, as was done in the case of X rays (Sec. 212).

Davisson and Germer of the Bell Telephone Laboratories carried out such an experiment in 1927. Their experimental setup was as indicated in Fig. 342. By means of an electron gun G a cathode beam was directed

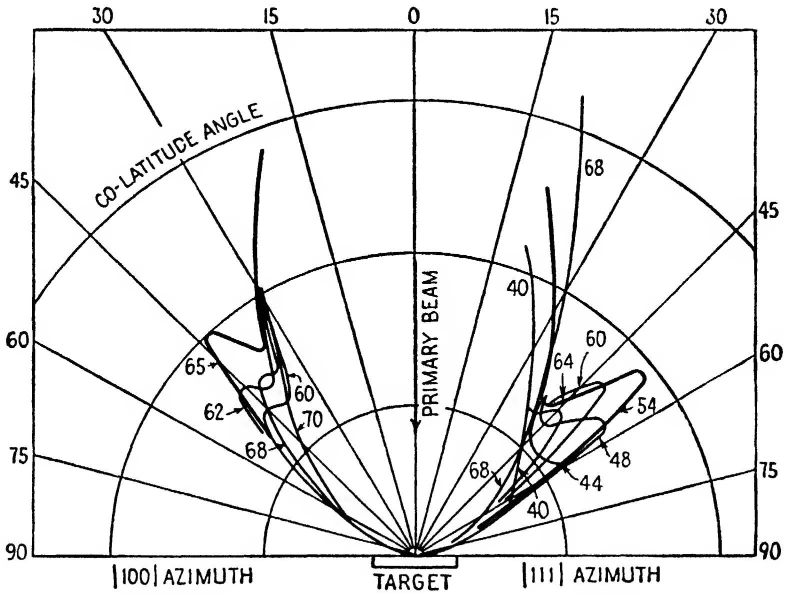


FIG. 343.—Scattering curves obtained by Davisson and Germer when using the experimental assembly shown in Fig. 342. Note the 54-volt and the 65-volt beams.

normally on to the face of a crystal of nickel. An electrode D connected to a suitable galvanometer served as a detector for the scattered electrons (or waves). By rotating this detector about the crystal as an axis, the value of the reflected energy could be observed as a function of the angle ϕ. Various runs were made using accelerating voltages ranging from 40 to 68; the results when plotted appeared as shown in Fig. 343. It will be observed that the electrons are scattered in all directions, and that the distribution depends upon the energy of the incident beam. Further, as the accelerating potential is increased, the most pronounced effect occurs when the potential difference is 54 volts and the angle 50°.

Davisson and Germer ascribed this peak to the reflection from the atomic planes existing in the crystal. They held that they were dealing

with an interference effect analogous to that which obtains in the case of X rays. In other words **the effect had all of the characteristics of constructive interference due to waves.** From the known atomic space, using the Bragg relation [Eq. (265)] and the data above given, the wave length proves to be 1.65 Å. On the basis of Eq. (273), the wave length value turns out to be 1.67 Å. Considering all of the experimental difficulties involved, this is a truly remarkable agreement. Thus the research of Davisson and Germer serves to provide strong evidence that the electron, at least, does exhibit wave properties. Both de Broglie and Davisson were later awarded the Nobel prize in recognition of their contributions in the domain of theoretical physics.

It is to Dr. G. P. Thomson, son of J. J. Thomson, that we owe another confirmation of the validity of the particle-wave concept. We have

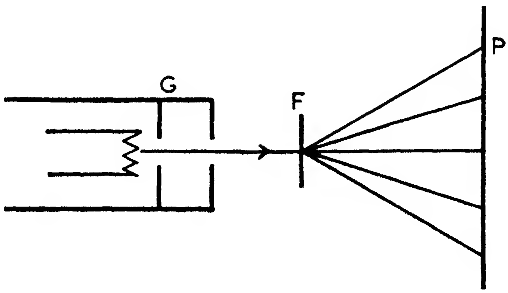


FIG. 344.—Experimental arrangement used by G. P. Thomson in obtaining diffraction photographs when an electron beam is passed through thin films.

already seen that diffraction patterns may be secured by passing X rays through crystals, thin films, and powders. In 1928 G. P. Thomson reported the results obtained when a beam of electrons was projected through thin films of various material. Metals such as aluminum, gold, and silver show a microcrystalline structure—crystals of microscopic size oriented at random. The general plan of Thomson's experimental assembly is sketched in Fig. 344. *G* represents an electron gun, *F* the metallic film, and *P* the photographic plate. The accelerating potential varied from 10,000 to 60,000 volts. The photographic records secured by Thomson showed interference patterns, as in the more familiar Laue diffraction records. A quantitative study of Thomson's photographs yields even more convincing evidence that electrons exhibit wave characteristics.

By using X rays to secure diffraction patterns of some given substance, such as aluminum, for example, one can readily determine the crystal spacing. By computing the probable equivalent electronic wave length by means of the de Broglie equation, and then measuring the details of

the electron diffraction pattern, one can compute the crystal spacing from the Thomson recordings. The agreement between the determinations is remarkably close. For instance, in the case of aluminum, the grating spacing as determined by X rays was 4.043, and by the electron method was 4.035. Various other metallic films were studied in a similar manner and the agreement was found to be equally close. Many diffraction recordings have been made in recent years. Figure 345 is a

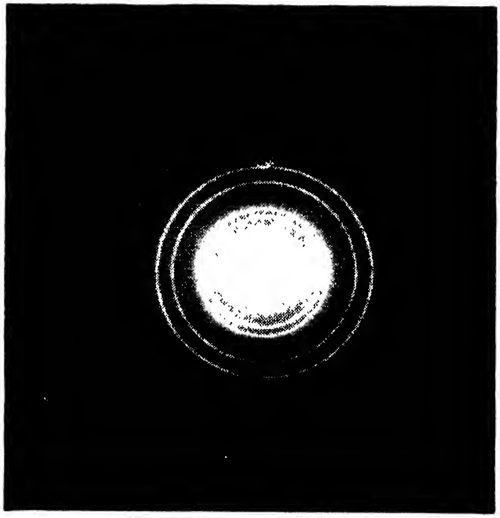


Fig. 345.—Electron diffraction photograph of gold foil. (*Courtesy of Professor Raymond Morgan, University of Maryland.*)

reproduction of such a photograph. The sharply defined interference pattern is obvious.

In summing up, then, it must be concluded that electrons exhibit dual characteristics, just as do photons. It should be noted, however, that **in a given experiment** such entities do not manifest **both** characteristics. We are still far from having a comprehensive well-established theory relating radiation and matter; but much progress has been made and, through further research, additional light will be shed on nature's intricate and beautiful interrelations. Who, for instance, will be the first to determine whether the positron or the proton also possesses dual characteristics?

PROBLEMS

1. The half life of Ra C is 19.7 min. What fraction of a given sample of this substance will be transformed in ten days?

2. How long a time would be required in order that only 10^{-6} of the original supply remained?

3. Could a knowledge of the half-life period of a radio-active substance be useful in its identification?

4. Compute the equivalent wave length of a 10,000-electron volt proton Express the result in Ångströms.

5. If the equivalent wave length of 100-volt electrons is 1.2 Å, what would be the crystal-grating space if the first order maximum reflection from a crystal occurred at 40°?

6. How much energy will be released when the nuclear reaction

$$Al^{27} + He^{4} \rightarrow Si^{30} + H^{1}$$

takes place?

7. The maximum orbital radius of deuterons being accelerated in a cyclotron is 50 cm. The magnetic field has a value of 5,000 gausses. What will be the kinetic energy of the deuterons as they emerge from the dees? Express the result in both ergs and mev. What would be the kinetic energy of α-particles under the same conditions?

CHAPTER XXIX

THERMIONIC TUBES AND THEIR USES

232. The Diode. The theory of thermionic emission was discussed in Sec. 188, and the control of free electrons by means of the anode and by a grid was considered in Secs. 189 and 192, respectively. These topics should be reviewed at this point.

We have seen that the two-electrode thermionic tube, or diode, exhibits unilateral conductivity, electrons passing from the filament (cathode) to the plate (anode). Extensive use is made of this property for the purpose of rectifying alternating current. For comparatively low

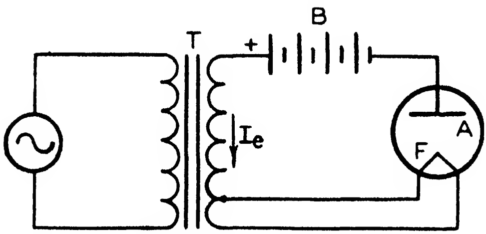


FIG. 346.—Diode used as a rectifying agent.

voltages two of the most common commercial units of this character are known as **"Rectigon"** and **"Tungar" rectifiers.** A typical circuit layout is shown diagrammatically in Fig. 346. The tube consists of a glass bulb filled with an inert gas, such as argon, and containing a short heavy tungsten filament F and a thick graphite plate serving as anode A. The filament is heated from a part of the secondary winding of the step down or autotransformer T. In the diagram, B represents a storage battery that is to be charged. If the filament F is heated, a copious supply of electrons will be liberated and, during the positive half cycle of the alternating emf, these free electrons will be attracted to the anode A. During the negative half cycle the electrons will be repelled by A, and hence the space current will be zero. In moving toward the plate A the electrons usually acquire a velocity of sufficient magnitude to ionize the gas present in the bulb, and thus augment the space current. The energy transfer through the tube will be in the nature of a pulsating unidirectional current that may be utilized to charge the secondary cells B, or for other purposes. Rectigon and Tungar rectifiers are made to deliver

currents up to 6 amp and potentials as high as 75 volts. The organization just described gives half-wave rectification, as indicated in Fig. 347.

Another widely used two-electrode rectifying device is known as the **mercury-arc rectifier.** A diagrammatic sketch of such a unit is shown in Fig. 348*a*. While this device actually has more than two electrodes, it is essentially the same as a two-electrode unit. It is so designed that it will handle relatively large currents and give full-wave rectification. The glass pear-shaped bulb has four connections sealed into the wall and terminating on the inside in graphite or iron electrodes. Mercury covers two of these electrodes, C and D. The two electrodes A and A' are

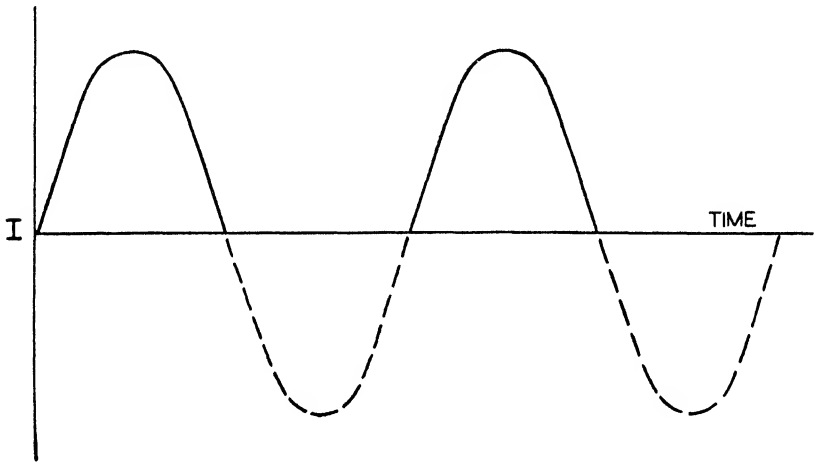


FIG. 347.—Half-wave rectification.

commonly called anodes, though during a part of the cycle they are actually negative. The bulb is thoroughly exhausted of air, with only mercury vapor remaining. When cold, the internal resistance of the tube is so high that current will not pass between either of the anodes and the cathode. If the mercury vapor is ionized, however, current will pass from either A' or A to C, depending on which of the anodes is positive. In order to start the rectifier, arrangements are provided for tilting the bulb so that a bridge of mercury is formed between the auxiliary electrode D, and C. Upon being returned to the normal vertical position, an arc between D and C is formed and some of the mercury is thus vaporized. The heated mercury at C liberates electrons which, by collision, ionize the mercury vapor in the region between the cathode and the two anodes A and A'. During one-half of each cycle one of the electrodes, A for instance, is positive and current will pass from it to C. It will thus be seen that the current in the load circuit is always in one direction, and

this is in effect a direct current. Unless some provision is made for maintaining the current while the emf at the terminals of the secondary of the transformer T passes through zero, the tube would cease to function. In order to obviate this a reactance coil L (sometimes called a sustaining coil), having a large inductive reactance and low ohmic resistance, is included in the load circuit. This reactance causes the current to lag somewhat behind the emf, with the result that there is still some current flowing from A to C, for instance, while it is beginning to flow from A' to C; thus the current in the load circuit at no time reaches zero. Once started, therefore, the ionizing process is continuous. The presence of the reactance also serves to reduce the emf ripple in the output.

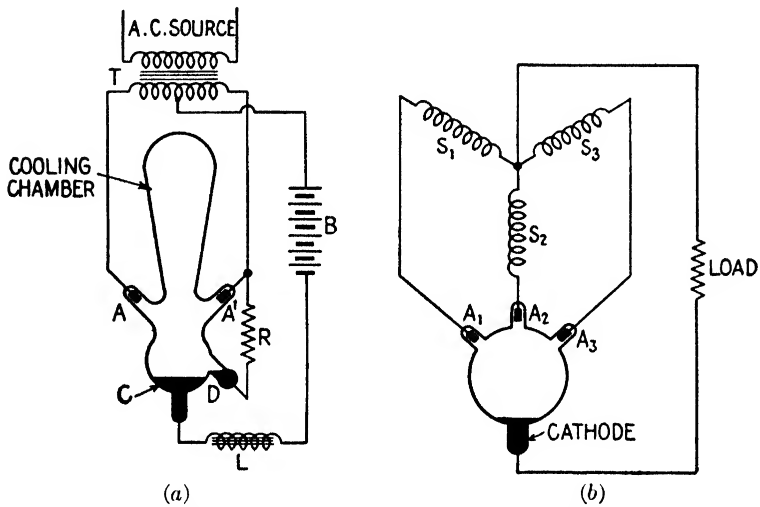


FIG. 348.—Mercury-arc rectifier. (a) Single-phase; (b) three-phase.

The unit just described would be called a single-phase rectifier. Rectification of the current in a three-phase circuit may also be accomplished. The essentials of a three-phase organization are shown schematically in Fig. 348b. In the sketch S_1, S_2, and S_3 are the secondaries of a three-phase transformer; A_1, A_2, and A_3 are anode terminals; and R the load resistance. In the three-phase rectifier no sustaining reactance is necessary.

During operation the voltage drop across a single-phase rectifier of the mercury type is about 14 volts. The over-all efficiency runs from 80 to 90 per cent, with a power factor approaching 0.9. Mercury-arc rectifiers are made in a variety of sizes and are used for charging storage batteries, for arc lighting purposes, and for supplying current to moving-picture equipment. Mercury-arc rectifiers have recently been

developed, using iron tanks instead of glass bulbs, that will deliver several thousand amperes.

A full-wave rectifying assembly that is widely used in the field of communication engineering makes use of two diodes and a "smoothing filter"; a diagram of the circuit employed is shown in Fig. 349. Frequently the components of two diodes are assembled in one glass inclosure, thus making the rectifier a one-tube unit, as shown in the sketch. The transformer T usually has two or more secondaries, the high-potential winding S_1, and the filament winding S_2. The filter is of the high-pass type, being so designed as to reduce the 120-cycle ripple to a negligible quantity. Such a unit is widely used to supply high potential (several hundred volts) for use in connection with audio amplifiers (Sec. 237) and radio receiving sets. In such cases the condensers, C_1 and C_2, would have capacitance

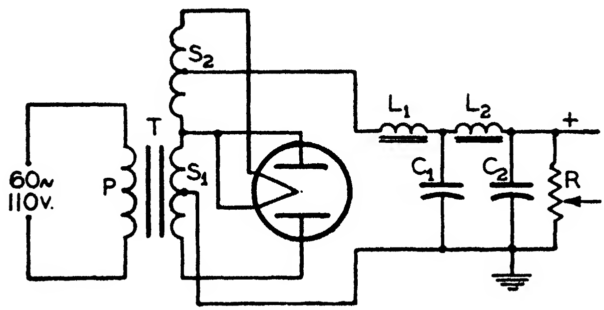


FIG. 349.—Full-wave rectifier, with filter.

values of the order of 8 μf each, and the inductances, L_1 and L_2, would have values ranging from 15 to 30 henrys. The resistor R, known as a "bleeder," serves to stabilize the output and also to discharge the condensers when the filter is disconnected: R has a value of the order of 35,000 ohms. The high-potential current supplied by such a rectifier is measured in milliamperes, though in the larger units (two-tube type) the rectified current may reach a value of an ampere or more.

Two types of diodes are used in rectifier assemblies of the character just described. One is a high-vacuum tube, which utilizes primary electrons exclusively. The other employs mercury vapor at low pressure, thus providing increased space current as a result of ionization. Mercury-vapor rectifier diodes are available in sizes ranging in output from a few milliamperes at a few hundred volts to units which will deliver as much as 5 amp at 20,000 volts. In this type of diode the tube voltage drop is largely independent of the load and is of the order of 15 volts. In the vacuum type diode the tube drop is higher and is a function of the load current. Vacuum diodes are made to handle anode voltages as

high as 50,000 and current capacities up to 8 amp. In units of this size, the anodes are usually water-cooled. Small vacuum diodes are also available. The type of rectifying tube to be employed, in any given case, depends upon the character of the load circuit. If the load is more or less constant, a vacuum diode may function satisfactorily; if, however, the load is constantly changing in magnitude, the mercury-vapor type of rectifier will usually meet the requirements better.

The assembly above described would be referred to as a single phase, full-wave rectifier. By the use of six diodes it is possible to effect full-wave, three-phase rectification. Such circuits are described in any standard work on communication.

233. Grid-controlled Rectifier Tube. For certain purposes the utility of the gas-filled rectifier tube, using mercury vapor or argon, can be increased considerably by the introduction of a control member between the cathode and anode. This component, which usually takes the form of a grid, performs a somewhat different function than the corresponding element in the triode (to be discussed shortly). In the grid-controlled rectifier the grid serves to **initiate** the passage of the space current, but once ionization begins the grid has no further effect on the space current, it functions as a triggering device. If the grid is held at a slight negative potential, the particular value depending on the anode voltage (0 to 5), no space current will pass; but if the negative grid potential is reduced, ionization will begin promptly, and the electronic current will continue until the anode potential again reaches zero. This effect is due to the fact that the positive ions form a space charge about the grid and thus neutralize its negative potential. By the use of such a tube, supplied by an alternating anode potential, it is possible to arrange conditions so that the current will pass for a definite fraction of a half cycle. Wide use is made of such a controlled rectifier. Among those applications may be mentioned their utilization in producing the sawtoothed wave form needed for actuating the sweep circuit in the cathode-ray oscillograph (Sec. 198), and their use in electric welding operations. When equipped with a metal housing, such a unit may deliver several hundred amperes for an interval of time measured in microseconds, the deionization time being of the order of 1,000 microseconds.

Grid-controlled rectifier tubes are referred to as **thyratrons, gas triodes,** or **grid-glow tubes,** depending upon the manufacturer, the first mentioned being the most common designation.

A tube that combines the control features of the thyratron and the high current capacity of the more common arc rectifier has been introduced by Slepian and Ludwig. This unit, known as the **ignitron,** has come into rather extended commercial use. Descriptions of the ignitron

are to be found in "Engineering Electronics" by D. G. Fink, and in "Electronics" by Millman and Seely.

234. Characteristics of a Triode. The fundamental theory of the three-electrode tube (triode) was considered in Sec. 192. We come now to a study of the operating characteristics of this very important device.

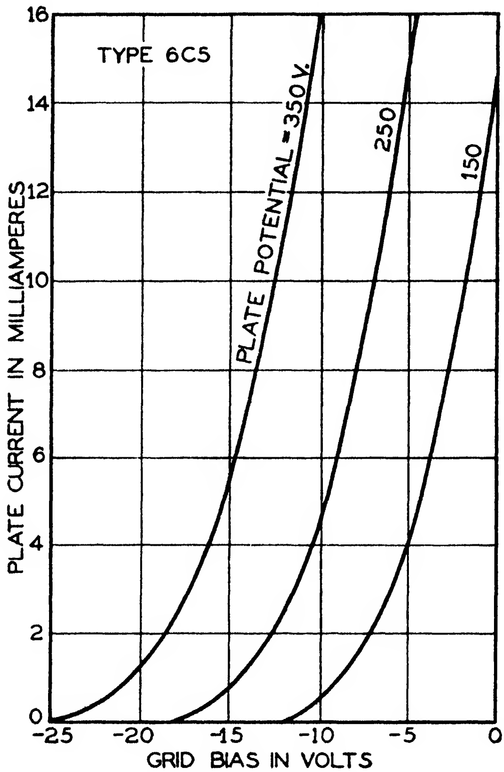


FIG. 350.—Graphs showing the relation between plate current and grid potential in the case of a representative triode (type 6C5).

It has been shown that the magnitude of the space or **plate current** for a given tube is given by the general relation

$$I_p = f(E_g, E_p, T_f), \tag{274}$$

where E_g is the grid potential with respect to the cathode, E_p the plate potential also with respect to the cathode, and T_f the temperature of the cathode. Since the cathode is usually held at a fixed temperature, we shall assume that this factor is constant. The plate current for a particular tube, then, becomes a function of the grid and plate potentials, I_p thus becoming the dependent variable, and E_g and E_p the independent varia-

bles. Much can be learned about the functioning of the triode from a study of the interrelation of the several factors mentioned above. Such a study involves what are called the **characteristics** of a tube; these distinguishing properties are usually represented in graphic form, the graphs being referred to as **characteristic curves.**

The relation between plate current and grid potential, when the anode potential is held constant, constitutes one of the important characteristics of the triode. A family of curves depicting this relation for a par-

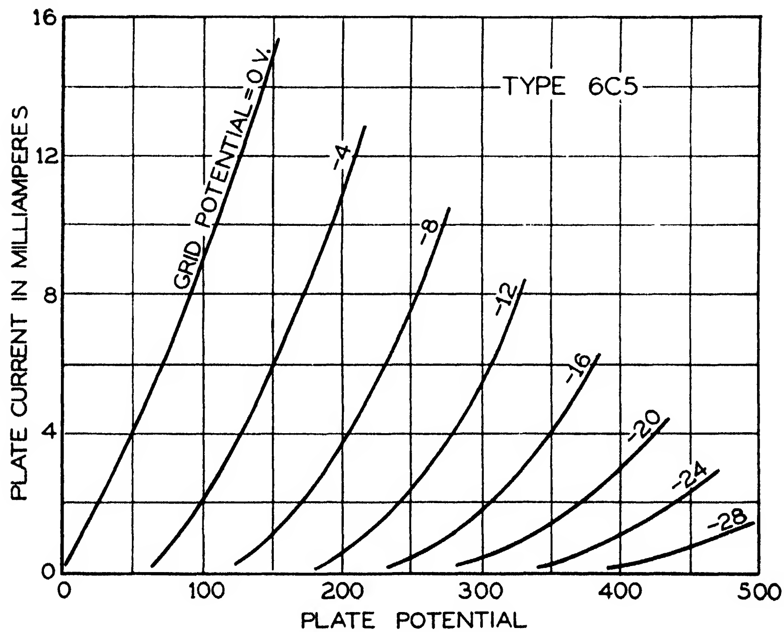


FIG. 351.—Curves showing the relation between plate current and plate potential for various biasing potentials in the case of a 6C5 triode.

ticular tube is given in Fig. 350. These curves bring out three facts: (1) at a certain negative grid potential the plate current becomes zero; (2) an increase in anode potential serves to shift the curves to the left without changing their slope; and (3) the graphs are approximately linear for a part of their length. If the grid had been made positive, the curves ultimately would have bent over toward the right, which would have indicated that **saturation had been reached.** However, with this particular tube, a positive grid potential would have resulted in excessive plate current and possible damage to the tube.

The relation between plate current and plate potential is also significant. Such a family of curves is sketched in Fig. 351. Here we see

what to expect in the way of plate current when both E_g and E_p are changed.

A third representation of the characteristics of a triode may be had by plotting grid potential against plate potential, as shown in Fig. 352. These graphs are drawn for various fixed values of plate current. They show the grid-plate relation which must obtain if the plate current is to be maintained at some given value.

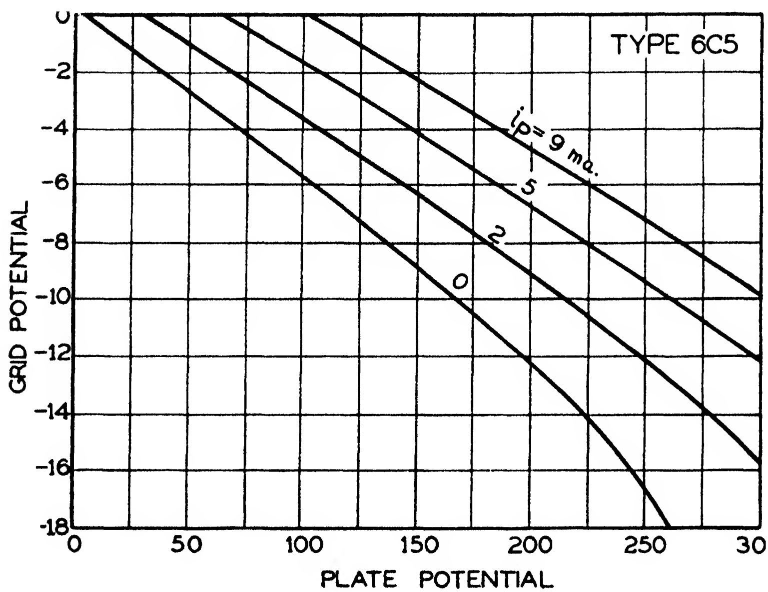


FIG. 352.—Graphs showing the relation between the grid potential and the plate potential of a typical triode (6C5).

235. Coefficients of a Triode. The most important coefficients of a triode are: (1) the *amplification factor* μ; (2) the *dynamic plate resistance* r_p; and (3) the *transconductance* g_m. (The last-mentioned factor is sometimes called mutual conductance.)

The **amplification factor** indicates the relative effectiveness of the plate and grid potentials in controlling the plate current. More specifically, it is defined by the relation

$$\mu = -\left[\frac{\partial E_p}{\partial E_g}\right]_{I_p \text{ constant}} \tag{275}$$

Partial derivatives are used because the factor I_p is held constant. The negative sign indicates that the plate potential and the grid potential must be changed in opposite senses if I_p is to be held constant. If a set

of E_p-E_g curves (Fig. 352) is available, the value of μ may be readily determined. The slope of any one of those curves will give the numerical value of μ for that particular value of plate current. From an examination of Fig. 353 it will be obvious that μ does not vary greatly. The value of this coefficient for any given tube will depend upon its structural details (such as the fineness of the grid mesh) and its relative position with respect to the cathode and the anode. Commercial triodes have μ values ranging from 1 to 100. For instance, if a tube had an amplification constant of 20, it would mean that 1 volt on the grid would produce the same change in plate current as a 20-volt change in the anode potential.

It is left for the student to discover how one might determine the value of μ from the I_p-E_g curves.

As the electrons move between the filament and the anode in the triode, they encounter what amounts to resistance. This opposition is due to the space charge and to any negative charge on the grid. If we were to divide the plate potential by the space current we should have what is known as the **static, or d-c, plate resistance.** However, this aspect of the picture does not concern us greatly. What **is** of importance is the so-called **dynamic plate resistance** of a tube, a relation giving the opposition offered by the plate circuit to a **changing** current. The formal definition of this quantity is

$$r_p = \left[\frac{\partial E_p}{\partial I_p}\right]_{E_g \text{ constant}}. \tag{276}$$

The value of this coefficient can be determined from any member of the I_p-E_p family of curves, as indicated in Fig. 351. It will be obvious that the magnitude of the dynamic plate resistance is a function of the plate current; hence when speaking of this coefficient one should always specify the corresponding anode current. The r_p curve shown in Fig. 353 should be examined in this connection.

The relation between grid potential and plate current is, in some respects, the most important of the three coefficients that we are considering. This factor (transconductance) has to do with the relation between the a-c components of grid potential and plate current, and is defined by the equation

$$g_m = \left[\frac{\partial I_p}{\partial E_g}\right]_{E_p \text{ constant}}. \tag{277}$$

Thus we see that this factor determines the change in plate current that is to be expected **per volt change in grid potential.** The value of this coefficient can be determined from any member of the I_p-E_g group of curves, as shown in Fig. 350. Inspection of the g_m curve (Fig. 353)

reveals that this factor is also a function of the plate current. The tube for which the curves are given has a transconductance of 2,000 μmhos, when the plate current is 8 ma. This means that the plate current will change 2 ma when the grid undergoes a change in potential of 1 volt. Triodes commonly used in radio receiving sets have transconductance values ranging from about 1,000 to 2,000 μmhos at normal plate current

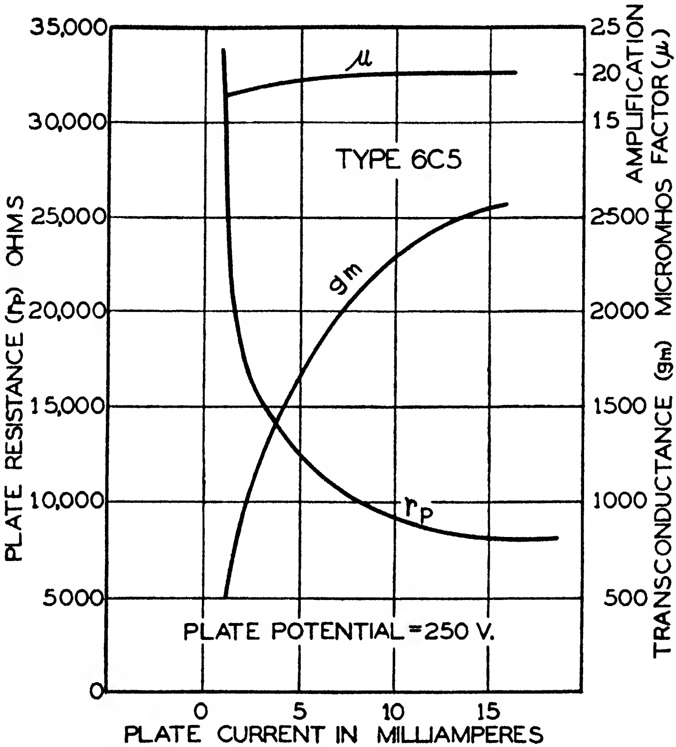


FIG. 353.—Curves showing the value of the several tube characteristics as a function of the plate current.

values. Tubes used as power amplifiers have values as high as 5,000 μmhos. The three coefficients above discussed are interrelated. A consideration of the defining equations [Eqs. (275) to (277)] will disclose that the relation

$$g_m = \frac{\mu}{r_p} \tag{278}$$

obtains. From this it is evident that the transconductance of a tube will be high if the amplification factor is high and the dynamic resistance low.

These three coefficients become basic considerations in connection

with the design of radio and other communication equipment. Methods of measuring these quantities experimentally are outlined in any standard laboratory manual on electronic tubes.

236. The Triode as an Amplifying Device. The triode is an exceedingly versatile device; it may, and does, perform many important functions—functions that, in recent years, have served to change the character of society and, indeed, the history of the human race. From the discussion of the preceding section it is evident that we have in the triode a device whereby a relatively small change in potential may be made to control a comparatively large amount of energy. It is this property of

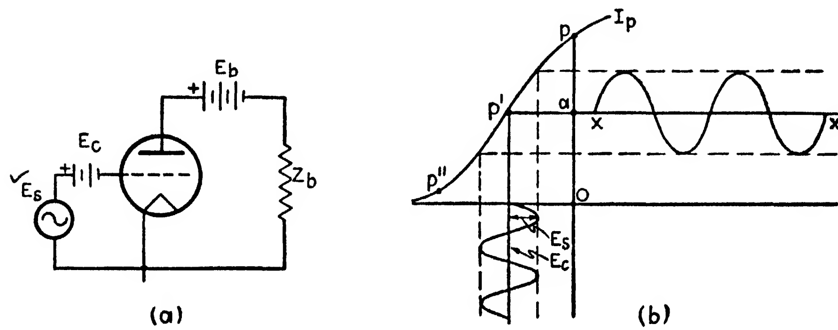


FIG. 354.—Triode used as an amplifying device.

a triode which is responsible for many of its important uses. Let us examine, somewhat in detail, how this may be accomplished.

The essentials of a single-tube amplifier are indicated in (*a*) of Fig. 354, and the potential and current relations are depicted in (*b*) of the same sketch. Z_b represents a load impedance, E_b the source of constant anode potential, and E_c is the so-called "biasing" potential applied to the grid. If E_c were made zero, and the signal voltage E_s also zero, the plate current would have a constant magnitude represented by *op*, the actual value being determined by the magnitude of E_b and the total ohmic resistance of the plate circuit. If, under such a condition (zero grid potential), a signal voltage were to be impressed on the grid, the control member would become **positive** during a part of its swing, with the result that the operating point *p* would traverse a part of the characteristic which is not straight; hence **the response in the plate circuit would not be linear.** Since, for most purposes, it is desirable that the variation in plate current shall be, so far as wave form is concerned, a replica of the signal voltage, it is necessary to so bias the tube that the quiescent operating point *p'* shall be roughly midway on the straight section of the tube characteristic. This is accomplished by applying a fixed negative potential E_c to the grid.

In the case of the 6C5 tube, for which certain data have previously been given, the proper biasing voltage is -8. With that potential on the grid the steady plate current (no impressed signal voltage) would be represented in magnitude by *oa*. If now a signal voltage, having the magnitude and wave form shown, is impressed on the grid—*i.e.*, between the grid and the cathode—the plate current will vary as indicated by the curved section xx'. In short, we will have a varying direct current in the load circuit of the tube. This current may be thought of as consisting of a **constant d-c component and an a-c component.** Under the circumstances just outlined, if the grid "swing" does not carry the operating point p' on to the nonlinear part of the characteristic, the alternating component of the plate will be a replica of the signal voltage. Due to the alternating component of the plate current, an alternating potential difference E_z will appear at the terminals of the load impedance Z_b, whose magnitude will be given by $Z_b I_{ac}$. This potential difference will, in general, be greater than the signal voltage.

The voltage amplification that may be secured will be given by the expression

$$A_v = \frac{I_{ac}Z_b}{E_s}, \tag{i}$$

where E_s is the signal voltage, *i.e.*, the potential difference applied between the grid and cathode. If the load impedance is a resistor, the above expression becomes

$$A_v = \frac{I_{ac}R_b}{E_s}. \tag{ii}$$

Since, as previously indicated, the drop across the load resistor would be

$$E_R = I_{ac}R_b, \tag{iii}$$

it follows that

$$A_v = \frac{E_R}{E_s}. \tag{iv}$$

Now the equivalent a-c circuit for a triode may be diagrammatically represented as depicted in Fig. 355. On this basis the alternating current in such a circuit is

$$I_{ac} = \frac{\mu E_s}{R_p + R_b}. \tag{v}$$

Combining the last two relations, we have

$$E_R = \frac{\mu E_s R_b}{R_p + R_b}. \tag{vi}$$

Substituting the value for E_R given in (vi) in (iv), there results

$$A_v = \frac{\mu R_b}{R_p + R_b}, \tag{279}$$

which gives the voltage amplification in terms of quantities which, in any given case, can be determined. In the last equation $R_b/(R_p + R_b)$ is always less than unity; hence the **amplification will be less than μ.** This equation is fundamental to amplifier design and operation, and should be carefully noted. If the amplifier is to be used for voice or music, the load impedance may consist of a pure resistance, the voice coil of a loud speaker, or the primary of a transformer.

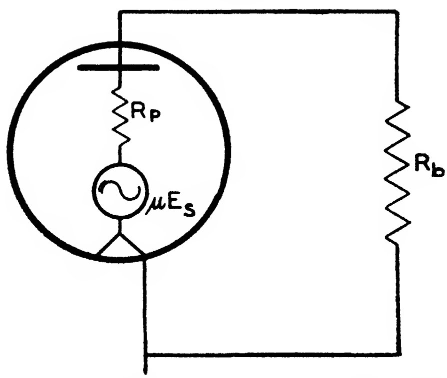


FIG. 355.—The equivalent a-c circuit in the case of a triode functioning as an amplifying agent.

237. Amplifiers. Amplifiers are classified either on the basis of their frequency range or on their mode of operation. Under proper conditions a triode will faithfully amplify any signal frequency ranging from 1 cycle/sec to many megacycles; it can also be made to amplify a direct potential. If the frequency range falls between, say, 35 and 15,000 cycles, the assembly is classified as an audio amplifier; if the range is expressed in kilocycles and megacycles, we would be dealing with radio amplifiers. Any triode will respond to this enormous gamut of frequencies, if and when it is associated with a circuit having suitable electrical characteristics.

On the basis of the mode of operation, amplifiers fall into three groups, the first of which is designated as class A. A **class A amplifier** is one that is so biased that the output wave form is a replica of the input potential wave. The conditions described in the preceding paragraph are those which would make that possible. Many audio amplifiers are of this type.

A second type of amplifier, known as **class B,** operates with a negative bias approximately equal to cutoff, *i.e.*, the tube is so biased that, with

no signal impressed on the grid, the operating point is at some such point as p'' (Fig. 354*b*). This means that when a sinusoidal signal voltage is impressed on the grid, the plate current will consist of a series of half-sine waves, similar to those which obtain in the case of a half-wave rectifier. Such amplifiers are used principally in radio transmitters, and occasionally in audio amplifiers. The class B amplifier is characterized by higher efficiency and higher output than the class A type.

A third form of amplifier is designated by the term **class C.** In this type, high output is the chief consideration. To attain this, the grid is biased considerably beyond cutoff, and a relatively high signal voltage is impressed on the grid. In operation the grid may swing positive, with

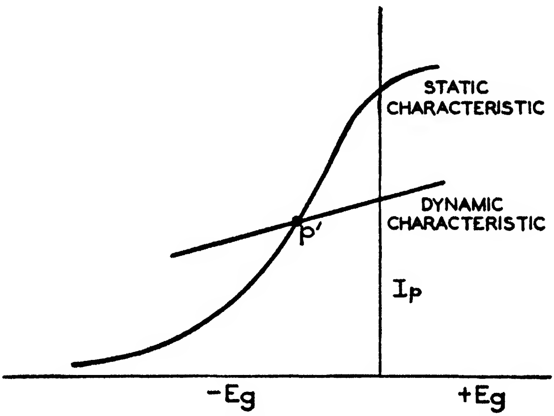


FIG. 356.—Relation of the dynamic characteristic of a triode to its static characteristic.

the result that the plate current may reach the saturation value. The efficiency and power output are higher than in the case of class B operation. Such amplifiers are employed only at radio frequencies.

It is to be noted that the essential difference between the several classes of amplifiers above described depends upon the magnitude of the biasing potential. It should also be observed that if the grid potential becomes positive during any part of the cycle **grid current will flow,** thus causing a distortion of the input wave form, which in the case of audio amplifiers is to be avoided.

The foregoing discussion has been based, for simplicity's sake, on the assumption that the static characteristic is identical with the operating, or dynamic characteristic. Actually, this is not the case except in one or two special instances.

A static characteristic curve, such as that shown in Fig. 354*b*, is made under the condition that there is **no external impedance** in series with the plate potential source E_b. When a tube is operating into a load (Fig.

354a), there will be a variable drop in potential in the anode circuit due to the presence of the load impedance. This will change, somewhat, the form of the characteristic curve. A curve taken under operating conditions is referred to as a **dynamic characteristic.**

As indicated above, the presence of impedance in the plate circuit will tend to change the slope of the characteristic curve; in general, the curve will be straightened and the slope decreased. Since any tube used as an amplifying device will operate into a load, as indicated in Fig. 354*a*, any increase in plate current will result in an increased drop over the load impedance; hence the potential actually applied to the plate will be decreased. As a result, the operating, or dynamic, characteristic will lie below the static curve, as shown by the slanted line in Fig. 356. Since

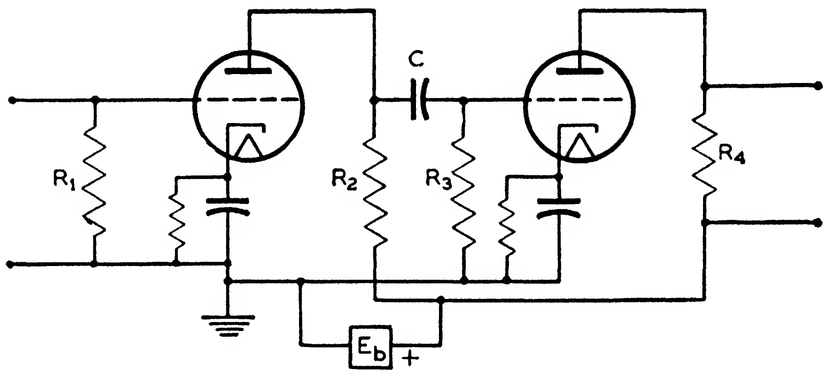


FIG. 357.—Resistance-coupled amplifier.

the load impedance may be, and often is, at least in part, a pure resistance, it may be said that the larger the load resistance the longer and straighter will be the dynamic characteristic curve. If, then, the load resistance R_b is large, there is less danger of distortion due to relatively large signal swings.

In the event that one triode tube with its associated circuit (single-stage amplifier) does not yield a sufficiently high amplification, one may arrange to couple successive tubes in such a manner that the output of one tube supplies the input potential difference to the succeeding stage. The three most common methods of coupling are (1) by means of a resistance and a capacitance; (2) by means of an inductive reactance and a capacitance; and (3) by means of a transformer. The circuit of a **resistance-capacitance coupled amplifier** is sketched in Fig. 357. The grid of the second stage utilizes the drop over R_2 as its impressed potential difference. This alternating potential is passed by the coupling condenser C. If necessary, R_4 may serve to supply a drop (potential

difference between its terminals) by which a third stage may be actuated. In the case of an audio amplifier, C will have a value of the order of 0.25 μf. The value of the coupling resistor R_2 will be in the region of 100,000 ohms.

A resistance-coupled amplifier operated as a class A unit gives a relatively high fidelity response.

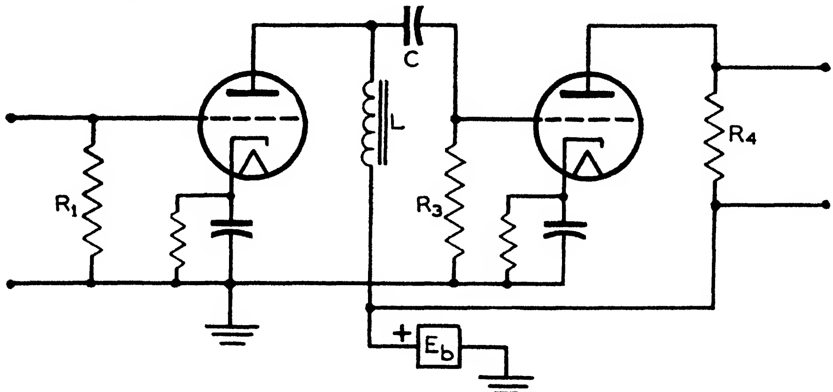


FIG. 358.—Reactance-coupled amplifier.

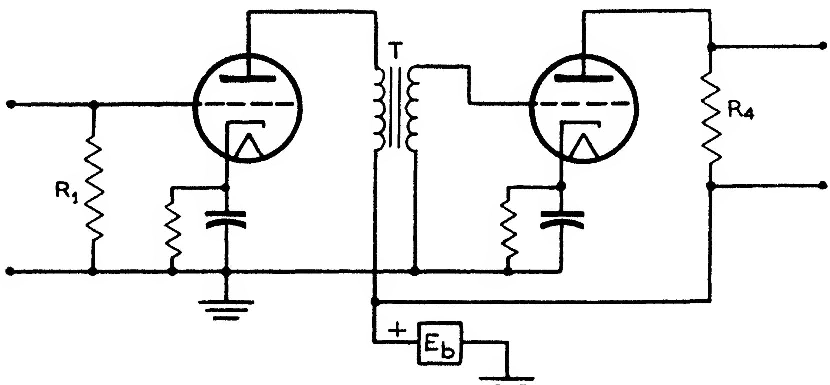


FIG. 359.—Transformer-coupled amplifier.

In the **reactance-capacitance coupled amplifier** (Fig. 358) an inductance replaces the coupling resistor R_2 of the resistance-coupled organization. In order to secure a suitable voltage drop across the resistor of the resistance-type coupling, a high value of resistance must be used. This results in a low plate current. This is avoided in inductively coupled amplifiers by making the magnitude of the inductance high and the resistance of the coil relatively low. Reference to a-c theory (Sec. 161) will show that the magnitude of an inductive reactance is a function of the

frequency. Hence this type of coupling will show some frequency discrimination. In practice, a compromise is usually made between the L and R values so that the distortion is not serious. In an audio amplifier of this type the coupling inductance would have a value of the order of 100 mh.

The circuit diagram for a **transformer-coupled amplifier** is seen in Fig. 359. This is a widely used type of coupling. The over-all gain per stage is relatively high, and, with well-made transformers, a high degree of fidelity can be attained. For mid-range frequencies the gain per stage is μn, where n is the step-up ratio of the coupling transformer.

Thus far in our discussion of amplifiers we have considered them only as a means of increasing the alternating potential difference (signal voltage) applied to the grid of the first tube. The triode may also be utilized for the purpose of **augmenting** the energy conveyed to the amplifying organization. Before discussing how this may be accomplished let us examine the power relations involved in the amplification process.

The power output of a triode operating in a circuit such as we have sketched in Fig. 354a would be

$$P = I_{ac}^2 R_b;$$

and, since we have already seen that

$$I_{ac} = \frac{\mu E_s}{R_p + R_b},$$

the above expression for power may be written

$$P = \frac{\mu^2 E_s^2 R_b}{(R_p + R_b)^2}. \tag{280}$$

If Eq. (280) is differentiated, and the result equated to zero, it will be found that the value of P will be a maximum when $R_p = R_b$. Under those circumstances

$$P = \frac{\mu^2 E_s}{4R_p}. \tag{281}$$

In this important relation, E_s is expressed in rms values.

To summarize, then, it may be said that, in order to secure maximum voltage amplification, the load resistance should be as high as feasible; and, to secure maximum power output, the resistance of the load or receiving circuit should be equal to the plate resistance. Our discussion might have been generalized—*i.e.*, we might have replaced resistance by impedance—but the form of the final result would have been the same. A general statement would be to the effect that the load impedance

should match the tube impedance in order to secure maximum transfer of power. It is to be observed, however, that a tube, when used as an amplifying device, does not create energy; the energy that it passes on comes from the local anode source of potential—from a battery or other d-c source. An ordinary transformer passes on only the energy supplied to it; **the electron tube releases energy from a local source.** This fact should be kept clearly in mind when dealing with electronic amplifiers.

There are many types of electron amplifiers, some of them using multielectrode tubes (Sec. 241). Each is designed for a particular purpose and to meet the general requirement of maximum power output with a minimum of distortion. We have confined our discussion to certain basic principles; details are available in any standard work on engineering electronics.

238. The Triode as a Generator of Alternating Currents. A second and highly important function that the triode is capable of performing is that of serving as a generator of alternating currents at frequencies ranging from a fraction of a cycle per second to several billion. When operating as an a-c generator it is commonly referred to as an **oscillator,** an oscillator being officially defined[1] as a "nonrotating device for producing alternating currents, the output frequency of which is determined by the characteristics of the device." It is **because of its amplifying characteristics** that a triode can function as an oscillator. Indeed a vacuum-tube oscillator may be thought of as a self-excited amplifier; instead of receiving an exciting potential difference from some outside source a part of its own output energy is fed back, under proper conditions, to the grid. Various circuits have been devised whereby this "feed back" may be brought about. We shall examine only one, by way of illustration; the diagram of a representative circuit is indicated in Fig. 360. It will be noted that the plate and grid circuits are inductively related, and that the plate inductance L_p is shunted by a capacitance C_p. If we think of the tube as an a-c generator, it will be seen that the L-C combination in the plate circuit forms a closed circuit that will have a definite oscillation frequency, and that this circuit is in parallel with the generator. (At this point the student should carefully reread Sec. 166.) E_c is a suitable source of biasing potential, and C is a so-called "blocking" condenser, its only function being to prevent a direct connection between the positive terminal of E_b and the grid. It is also to be noted that the source of plate potential E_b is connected **in parallel** with the tube and the plate circuit, instead of in series as in the amplifiers previously examined. (Any amplifier could be so connected, and often is.) With this type of E_b connection, **no direct current flows through the**

[1] *Standards, IRE.*

L-C circuit. L is an inductance the impedance of which is of such a magnitude that it will not pass an appreciable amount of alternating current of the frequency which is being generated by the tube; it is referred to as the "plate choke coil."

Having examined the circuit as such, the next question that presents itself is: How are the electrical oscillations started, and once started how are they maintained?

To find an answer to that question, let us suppose that the filament is heated and that thermionic emission is taking place; but let the connection to E_b be open. Suppose now that we close this circuit. Space current I_p will at once begin to flow, and **a potential pulse will also have**

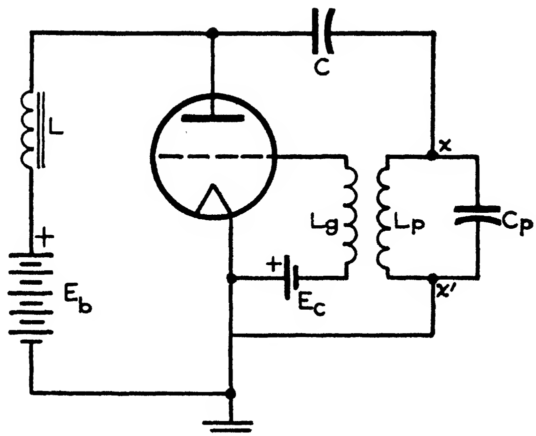


FIG. 360.—The triode as an a-c generator.

been impressed on plate L-C circuit. This potential pulse will initiate a minute oscillating current in the L_p-C_p circuit. The situation is somewhat analogous to the case of a pendulum when slightly touched. The pendulum will be set into mechanical oscillation in its own natural period of vibration. So, too, will the circuit being considered; it will oscillate electrically at a frequency determined chiefly by the constants L_p and C_p, though the amplitude of the oscillations will be exceedingly small and they will quickly die out unless additional energy is supplied from some outside source.

Now, since L_p and L_g are inductively related or "coupled," the small alternating current in the circuit L_pC_p will, by induction, give rise to an alternating emf in L_g. Thus an alternating potential, having a frequency determined by the constants of the parallel resonant circuit, will be impressed on the grid of the triode. This alternating grid potential will, as we have seen, cause corresponding variations in the direct

current through the tube. These current variations will give rise to a variable potential difference between the points x and x'. This will result in an augmented oscillating current in the L-C circuit, which, in turn, will give rise to a wider swing of the induced grid potential. Thus it is seen that the process is cumulative, the additional energy required being supplied by the d-c source E_b. The growth of the oscillating current in the parallel resonant circuit is subject to certain well-defined limitations, which we will examine shortly. But for the moment let us glance at the **phase relations** that must obtain **if oscillations are to start.**

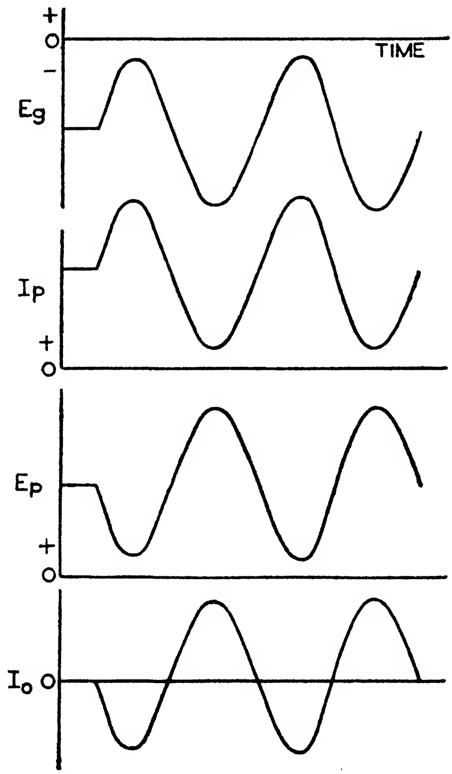


FIG. 361.—Phase relations in the case of a triode when functioning as an oscillator.

Keeping in mind that the tube, in effect, acts as a generator that is developing an alternating potential equivalent to μE_g, and also noting that the grid of the tube has a fixed negative bias, it will be seen that the alternating component of the plate potential will differ in phase by 180° from the alternating component of the plate current. This a-c potential, which appears between the points x and x', constitutes the voltage across the primary L_p of an air-core transformer of which L_g is the secondary. We have seen (Sec. 146) that the voltage induced in the secondary of a transformer differs in phase by 180° from that in the primary. Hence the alternating potential induced in the grid circuit will be in such a direction as to tend to sustain the oscillations, **if and when the coils L_p and L_g bear the proper inductive relation to one another.** This condition obtains when the windings in the two coils have their turns in the same direction. The phase relations above referred to are graphically represented in Fig. 361; the chart should be carefully studied.

If the energy supplied to the anode circuit by E_b is **less** than that necessary to compensate for the I^2R loss of the L-C circuit, the oscillations will **decrease** in amplitude and eventually cease. If, however, the

energy supplied is equal to the loss above mentioned, the oscillations will just be sustained. In the event that energy is abstracted from the L-C circuit by means of an associated load circuit, additional energy must be supplied by E_b. The degree of excitation can be changed by adjusting the coupling between the plate and the grid circuits. In order

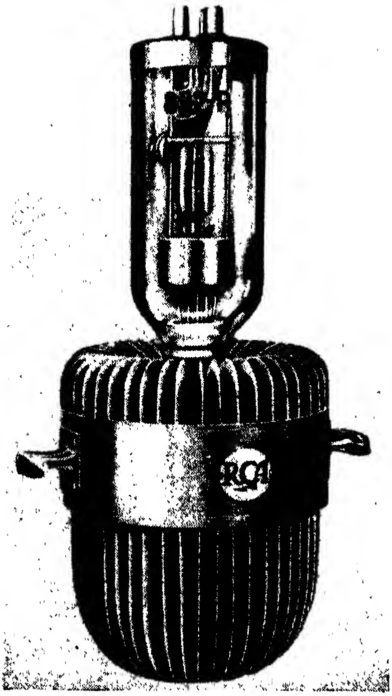


FIG. 362.—Typical air-cooled triode designed for use as an r-f power oscillator or amplifier. (*Courtesy of Radio Corporation of America.*)

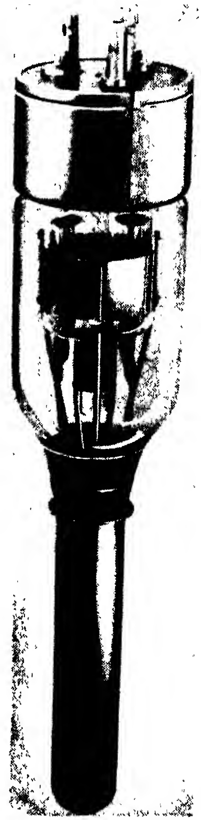


FIG. 363.—Water-cooled triode for use as an r-f power oscillator or amplifier. (*Courtesy of Western Electric Co.*)

to accomplish this, if the case be one of electromagnetic coupling, the relative physical position of the grid and plate coils may be changed, or their turn ratio modified. If the coupling is electrostatic, the value of the coupling capacitance is varied.

The maximum oscillating current amplitude will be limited by two conditions which depend upon the characteristics of the particular tube

being used: (1) the anode current cannot have a value less than zero, and (2) the saturation value of the plate current will limit its maximum. There is also an additional factor that must be considered in determining the maximum energy that any given tube can deliver to a load circuit. This factor is the **energy-dissipating ability of the anode.** The bombardment of the plate by the electrons produces heat, and this thermal energy must be disposed of or the tube will be electrically and mechanically wrecked. In tubes rated at a few watts oscillating-current output,

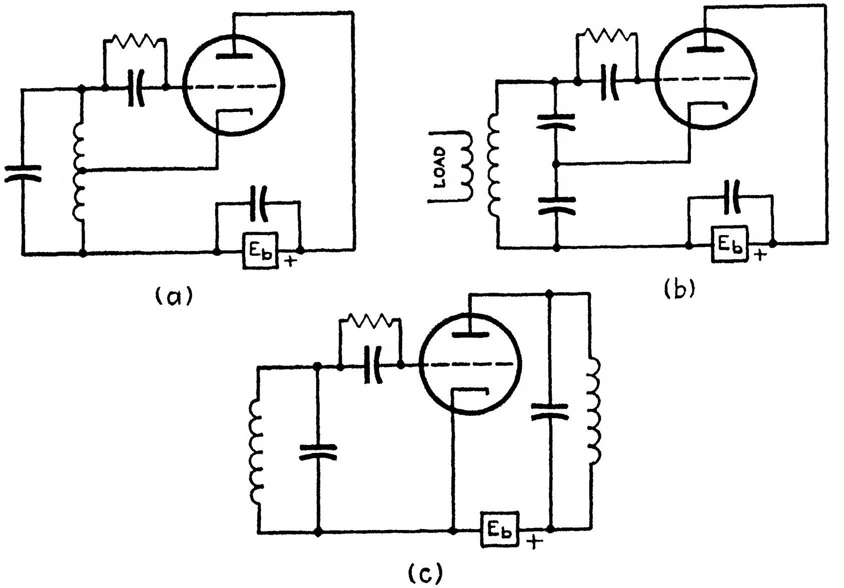


FIG. 364.—Circuits most commonly used for the production of electric oscillations (*a*) Hartley; (*b*) Colpitts; (*c*) tuned-plate tuned-grid.

simple radiation from the plate is depended upon to accomplish this end. In the case of tubes rated in kilowatts, some form of air or water cooling is provided. An air-cooled triode is shown in Fig. 362, and a water-cooled unit is illustrated in Fig. 363. **The power rating of a tube is based upon its ability to dissipate thermal energy at the plate.** For example, a 50-watt tube will dissipate 50 watts as thermal energy at the anode. Triodes used as oscillators operate at efficiencies ranging from 50 to 75 per cent.

Various circuits have been devised by the use of which a triode may be made to serve as an a-c generator. Several commonly used circuits are sketched in Fig. 364. These circuits illustrate various ways of feeding back to the grid a certain part of the output energy in the proper phase relation.

In the tuned-plate tuned-grid oscillating circuit *c*, the energy transfer from the plate to the grid takes place electrostatically, the interelectrode capacitance acting as the condenser. All of these circuits are equally efficient. The particular circuit to be used is determined, in part at least, by the nature of the electrical load.

In dealing with the triode as an a-c generator, the question of frequency stability naturally arises. In general terms, the magnitude of the inductance and capacitance of the plate circuit determines the frequency of the electrical oscillations being developed. (This circuit is frequently referred to as the "tank" circuit.) It may however be

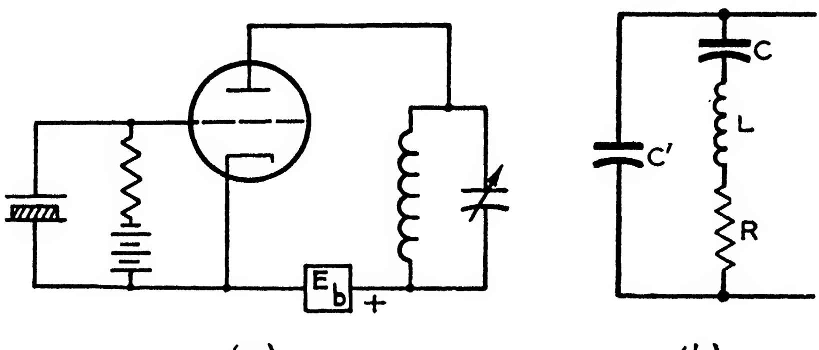


FIG. 365.—Use of a quartz plate for the control of the frequency in an oscillating circuit. (*a*) Circuit commonly employed; (*b*) the circuit which is equivalent to that shown in (*a*).

shown[1] that the frequency of a tuned-plate oscillator is given by the expression

$$f = \frac{1}{2\pi}\sqrt{\frac{R_o + R_p}{L_oC_oR_p}}, \tag{282}$$

where R_o is the resistance of the oscillating (tank) circuit, L_o its inductance, and R_p the tube's plate resistance. It will be recalled [Eq. (276) and Fig. 353] that the plate resistance is not a fixed quantity. Any variation in the emission or the anode-supply voltage will cause slight changes in R_p, and hence, because of relations indicated by Eq. (282), will result in a change in the frequency. If a high degree of frequency stability is desired, as is usually the case, recourse is had to the use of piezoelectric control (Sec. 110).

The circuit assembly in which a quartz crystal may be used for the control of frequency is sketched in Fig. 365*a*. The mechanical char-

[1] A clear discussion of this topic, in mathematical form, is to be found in "Principles of Radio Engineering," by R. S. Glasgow, p. 264.

acteristics of a piezoelectric crystal are equivalent to certain electrical quantities, as indicated in Fig. 365*b*. C' represents the capacitance of the two plates of the crystal holder. The series-resonant circuit elements C, L, and R represent the corresponding mechanical quantities of mass, compliance, and friction. When the plate circuit is adjusted to approximate resonance with the natural frequency of the crystal, oscillations take place **at the crystal frequency.** The energy required to maintain the oscillations is fed back electrostatically through the capacitance existing between the plate and the grid electrodes. The output of such an oscillator unit is commonly augmented by one or more stages of amplification, as we shall see later. A very high degree of frequency stabilization may be secured by such a combination.

239. The Triode as a Modulator. Energy in the form of a h-f alternating current may be conveyed to an antenna system and from thence radiated into space; or energy may be impressed upon a physical line and thereby guided to some distant point. In either event, if intelligence is to be transmitted, some means must be provided whereby the h-f current (the so-called carrier wave) can be modified in conformity with a system of signals, or by sound waves. Such a process of modification is known as **modulation,** and we shall see that a triode tube may be made to accomplish this end.

There are three methods by which a modification of the carrier wave may be effected; these are amplitude, phase, and frequency modulation. Our limitations will permit discussion of the first type only, which, incidentally, is the most widely used, at least for sound transmission.

Amplitude modulation consists in modifying the amplitude of the carrier wave in conformity with the character of the signal. The process of amplitude modulation may be graphically illustrated as shown in Fig. 366. The upper trace *a* represents an unmodulated carrier wave whose frequency is, say, 10^6 cycles/sec. If by some means we could modify the amplitude of the "carrier" by means of, for instance, a sinusoidal 1,000-cycle signaling wave *b*, the modulated wave would take the form shown in *c*. The curved line would be spoken of as the **envelope** of the modulated carrier wave and would correspond to the shape of the modulating signal.

The degree to which the carrier wave is modulated is given by

$$M = \frac{A'}{A}, \tag{283}$$

A and A' having the significance indicated in Fig. 366*c*. For purposes of measurement, the degree of modulation is given by the relation

$$\text{Percentage modulation} = \frac{I_{max} - I_{min}}{I_{max}} \times 100. \tag{284}$$

A case where the modulation is 100 per cent is illustrated by the trace marked *d*. The fundamental principles involved in amplitude modulation can be outlined as follows.

Assuming, by way of illustration, that the carrier is a simple sinu-

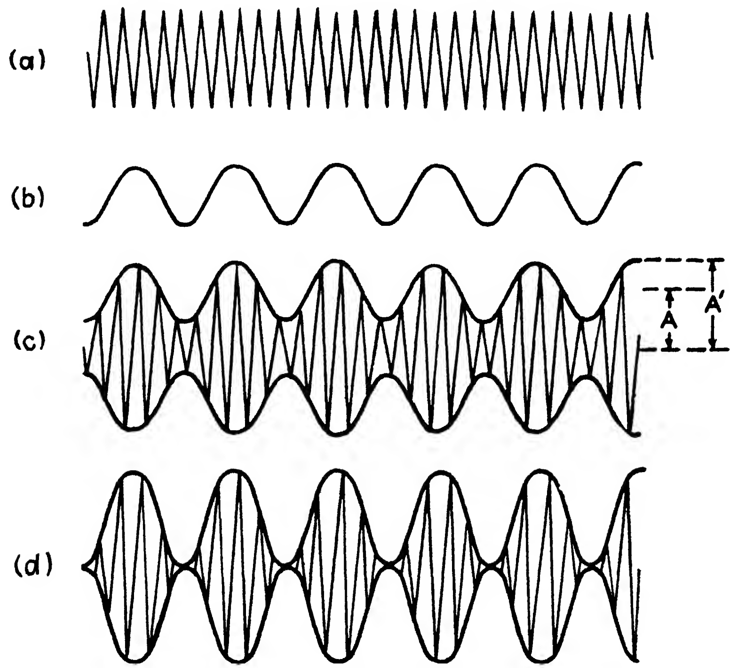


FIG. 366.—Amplitude modulation. (*a*) Carrier wave; (*b*) modulating wave; (*c*) partial modulation; (*d*) 100 per cent modulation.

soidal wave, the instantaneous value of that current wave would be

$$i = I_m \sin \omega_c t, \tag{i}$$

where the subscript c refers to the carrier current. If the amplitude of such a current wave is caused to vary in a sinusoidal manner and at a frequency less than that of the carrier, the maximum current amplitude would be given by

$$I_m = I_c(1 + m \sin \omega_s t), \tag{ii}$$

where I_c is the carrier maximum; ω_s the signal radian frequency; and m the modulation factor, which may be anywhere between zero and unity.

If the value of I_m given by (ii) is substituted in (i) we have

$$i = I_c(1 + m \sin \omega_s t) \sin \omega_c t. \qquad \text{(iii)}$$

Expansion of (iii) by trigonometric methods leads to the interesting and significant relation

$$i = I_c\left[\sin \omega_c t + \frac{m}{2} \cos (\omega_c - \omega_s)t - \frac{m}{2} \cos (\omega_c + \omega_s)t\right]. \qquad (285)$$

An examination of the above equation indicates that three current waves **coexist,** one having the original carrier frequency, one having a frequency of $f_c - f_s$, and another having a frequency of $f_c + f_s$, where the subscripts indicate carrier and signal frequencies, respectively. Returning to the numerical values of f_c and f_s, as cited on page 530, it may be seen that the three frequencies would be 1,000,000, 1,001,000 and 909,000; the two last frequencies being referred to as **side frequencies.** If the sound to be transmitted is composite, as is usually the case, each component will constitute a signal frequency and each such component will give rise to a pair of side frequencies. Thus there will exist two groups of side frequencies, which are commonly referred to as the **side bands.** It is thus evident that the transmission and reception circuits must be designed to respond simultaneously to several different frequencies, and that the range of frequencies thus involved is twice as great as the corresponding range of frequencies in the original signal wave. For instance, if music having a frequency range lying between 100 and 10,000 is being transmitted, the required transmission band would have a width of 8,000 cycles. In the case of speech the approximate frequency range is from 200 to 3,000 cycles. The transmission band would be 5,600 cycles. If the transmission and reception circuits are not capable of transmitting the required band of frequencies, distortion occurs.

There are a number of methods by which amplitude modulation may be effected. Only the plan most widely used will be touched upon here. For the transmission of voice or music, this method involves a low-power class C stage of r-f amplification whose plate circuit is coupled to the output of an audio amplifier, operated class A or B. The minimum elements of such a circuit assembly are sketched in Fig. 367. The triode V_1, and its associated grid and plate circuits, functions as an intermediate r-f amplifier stage. The triode V_2 is the output stage of an audio-power amplifier. It will be noted that the anode circuit of V_1 contains **two sources of emf in series**—one the usual d-c supply E_b, and the other the a-f emf developed in the secondary of the transformer T. If, then, an audio signal is impressed on the grid of the

modulator V_2, the plate potential of V_1 will be changed (at audio frequency) from a value less than E_b to a value greater than E_b. As a result, the output of the r-f amplifier will vary in a corresponding manner, and the carrier wave will be modulated. It should be understood that it is necessary, in actual practice, to modify the above indicated circuit in a number of important details, but the essentials of so-called "plate modulation" are as outlined.

In passing, it should be mentioned that another modulation technique has recently come into limited use. Reference is made to the so-called

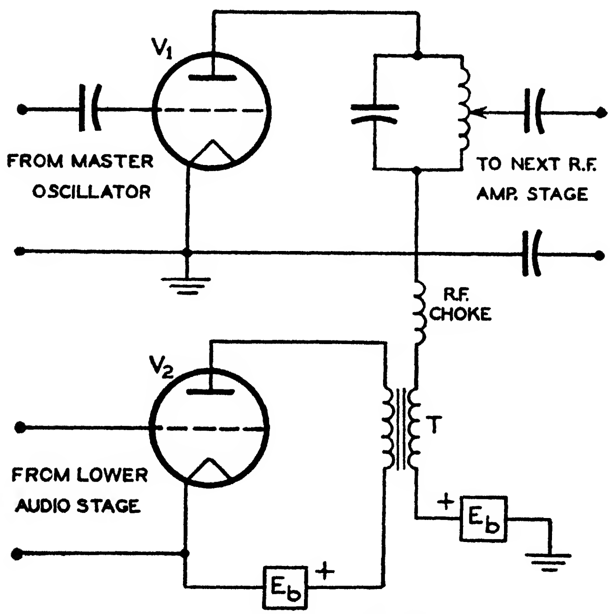


FIG. 367.—A common method of effecting amplitude modulation.

frequency-modulation system. In the system which has just been described (amplitude modulation), the frequency is held constant while the amplitude of the carrier wave is modified. In the system to which we are now referring, the amplitude of the carrier is held constant and the **frequency** is caused to change in response to an audio signal. There are several methods by which frequency modulation may be brought about, the fundamental principles having been known for many years.

In Sec. 238 it was shown that a triode may be caused to produce an alternating current of any desired frequency, and that the frequency thus developed is determined by the values given to the inductance and the capacitance in the plate circuit (L_p and C_p, Fig. 360). It follows, therefore, that any change in either the inductive or the capacitive

reactance of the tank (plate) circuit will result in a change in the frequency of the alternating current developed by the oscillator. In a system of frequency modulation developed by Dr. E. H. Armstrong, and now being utilized commercially to some extent, the **total** reactance of the tank circuit is caused to change at audio frequency. This variation in the total reactance of the oscillator plate circuit is accomplished by the use of a multigrid tube, such as the 6L7, and its associated circuits. Thus the carrier frequency is made to deviate from a mean value by an amount called the **deviation frequency.** The **rate** at which this change in the carrier frequency occurs is a function of the modulation frequency, and the **magnitude** of the frequency deviation is controlled by the strength of the microphone signal.

In order to demodulate an f-m carrier wave, and thus recover the signal, it becomes necessary first to convert the f-m carrier current into an a-m current, and then to demodulate the a-f component by any one of the methods described in the following section.

Because of the nature of the process, a relatively wide band of frequencies is required for each channel operated on a frequency-modulation basis. For this and other reasons, frequencies of the order of several hundred megacycles are employed. The chief advantage claimed for this type of modulation is that the signal-to-noise ratio is better than when amplitude modulation is used. A part of this gain is probably due to the use of a h-f carrier, rather than to the particular type of modulation.

A description of the details of the process of frequency modulation and the demodulation of such signals is outside the scope of this text. The interested reader will find a brief account of this type of modulation in "Basic Radio," Chap. XXXIII, by J. B. Hoag. A thorough treatment of the subject is to be found in a paper by W. L. Everitt in *Electrical Engineering* for November, 1946.

240. Demodulation or Detection. If we assume that a modulated r-f carrier current set up in some form of receptor circuit has been, by r-f amplification, raised to a suitable energy level, it then becomes necessary to **demodulate** the carrier in order that the original intelligence may be recovered. In other words, we must in some way separate the signal wave from the carrier. This is usually accomplished by means of some agent whose electrical response is nonlinear, *i.e.*, by some device that shows nonsymmetrical conductivity. There are various such agencies; only the two most commonly used will be mentioned.

One method (called plate detection) involves the triode, the circuit of which is sketched in Fig. 368. It will be evidence that this circuit constitutes a simple amplifier having a resistive load. The grid bias

and anode voltage are so chosen that the tube operates on a nonlinear portion of its static characteristic, usually at the lower knee of the curve. (The grid bias is secured from the d-c drop over the resistor R_1. The condenser C_1 serves to by-pass both r-f and a-f current around R_1.) As a result, the dynamic characteristic will be of such a nature that the upward swing of the r-f carrier wave in the plate circuit will be more pronounced than the lower swing, as indicated in Fig. 369. It will thus be seen that the triode functions as a partial rectifier. [The r-f component, in the plate circuit, is by-passed by the condenser C_2 (0.001 to 0.002 μf)]. Since the device causes a difference in amplitude between the upper and lower swings, the **average** amplitude **is not zero,** and hence a current that has the frequency and wave form of the signal

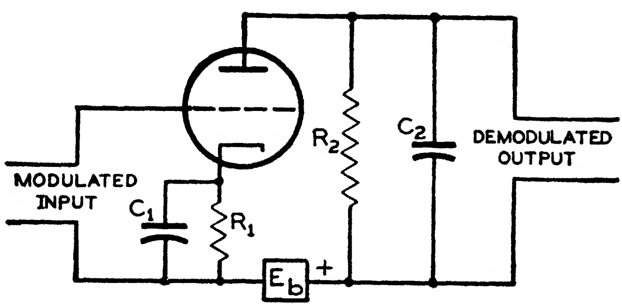


FIG. 368.—Demodulation by means of a triode.

wave will exist in the plate circuit. This a-f current will flow through the resistor R_2 (50,000 to 100,000 ohms), and the drop across the resistor becomes the recovered signal wave, which, in turn, may be made to undergo sufficient audio amplification to operate a loud-speaker, or other signal-translating device. From the nature of the circuit, this method of recovering the signal wave not only effects demodulation but also serves to amplify somewhat the incoming signal. A more comprehensive discussion of plate detection will be found in "Principles of Radio Engineering" by R. S. Glasgow, p. 345.

A somewhat less complicated method of demodulation makes use of a simple diode as a rectifying device. A common circuit layout is shown in Fig. 370. The diode as a rectifying agent has already been discussed (Sec. 232); in this instance a diode is used to rectify an h-f current. The varying, but rectified, current flowing through the series resistor R will have the wave form and frequency of the signal wave, and thus demodulation is accomplished. The drop over the resistor R (0.5 to 1.0 megohm) becomes the signal output, and this is amplified to the desired level, as in the case of the plate-circuit demodulator previously described. The

condenser C (order of 100 $\mu\mu$f) serves to conduct any unrectified carrier current around the high resistance R and thus prevent its transfer to the associated audio-amplifier circuit. Both methods of demodulation described above have their advantages and disadvantages; in any given

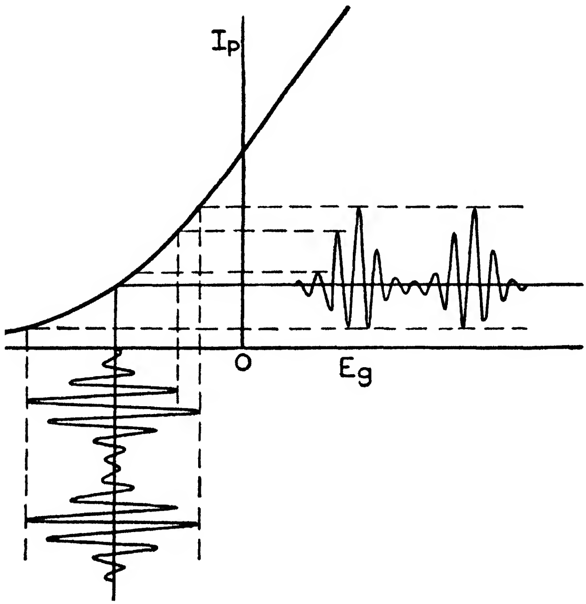


FIG. 369.—Showing the manner in which demodulation is effected by the use of the circuit shown in Fig. 368.

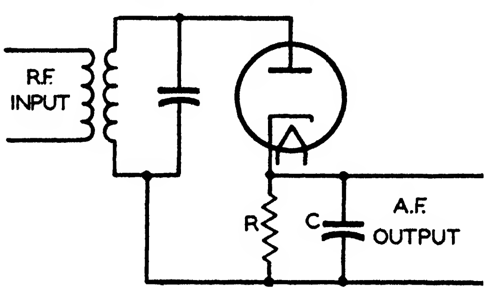


FIG. 370.—Demodulation by means of a diode.

instance that method is used which will best meet the particular requirements of the case in hand.

241. Multielectrode Tubes. A certain amount of capacitance exists between the several electrodes of a triode; each electrode acts as one element of a condenser. Therefore there is plate-to-grid, plate-to-cathode, the grid-to-cathode capacitance, the first mentioned being the

greatest (of the order of 1 to 15 $\mu\mu$f), and therefore the most significant. In a tuned-plate, tuned-grid, r-f amplifier, the interelectrode capacitance of the grid and plate may be great enough to bring about electrostatic coupling between the plate and the grid circuits; and thus give rise to oscillations. This is highly undesirable, and external or internal means must be provided to prevent or suppress such an electrical reaction. In triodes, used either as audio or r-f amplifiers, the grid-to-plate capacitance, by permitting alternating current to pass between these two electrodes, has the effect of reducing the tube impedance to a low value, thereby changing the characteristics of the tube for different frequencies The net result of these effects is to limit the maximum ampli-

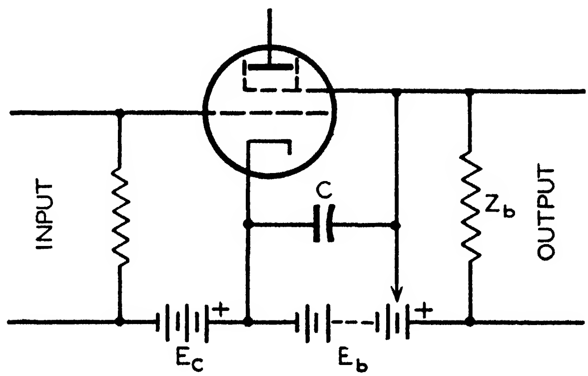


FIG. 371.—Functioning of a four-electrode tube, the tetrode.

fication obtainable. By introducing a **fourth** electrode into the tube, the grid-to-plate capacitance can be greatly reduced, and thus the performance of the tube be materially improved, particularly when used in r-f amplifying circuits. Such a tube is called a **tetrode.** The fourth electrode commonly takes the form of a shielding grid structure which surrounds the plate. The relative positions of the several elements are sketched in Fig. 371. The by-pass condenser C enhances the shielding effect due to the so-called **screen grid.** The presence of the screen grid reduces the interelectrode capacitance from several micromicrofarads to a small fraction of a micromicrofarad. The screen grid is operated at a positive potential with respect to the cathode, and at a value of about half that of the anode. The screen grid, therefore, attracts electrons toward itself, though most of them eventually reach the plate. However, the screen grid acts as an accelerating electrode, and thus the potential of the plate is not as effective in determining the space current as it is in the simple triode. In fact, over most of its characteristic, the plate current is largely independent of the magnitude of the plate potential.

But the screen grid does not materially affect the action of the control grid. Hence the amplification constant of such a tube is high. In a type 36 tube, for instance, μ has a value of nearly 600.

A family of characteristic curves for the type 36 tube is shown in Fig. 372. Several features of these curves are to be noted. It will be seen that the plate current is nearly independent of the plate potential, thus indicating a high dynamic plate resistance. The dip in the curves at low plate-potential values (below 100) is due to secondary emission. In the triode, any secondary electrons liberated at the plate are soon recaptured by the plate itself, but in the tetrode some of the secondary

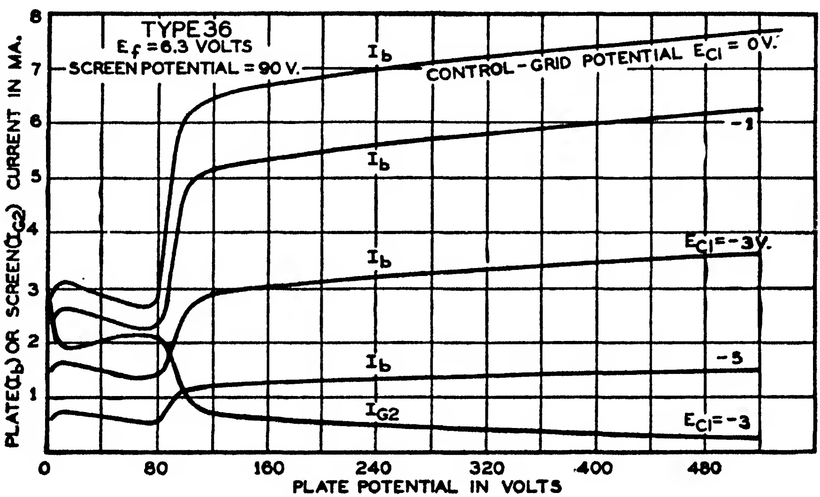


FIG. 372.—Operating characteristics of a tetrode.

electrons are collected by the screen grid, and thus there is a certain **inverse** electron current which has the effect of reducing the total plate current; hence the dip in the characteristic at low plate voltages. A tetrode of this type is seldom operated at plate potentials below 100 volts. Under any circumstances, there is a relatively small screen-grid current.

It has been found possible to suppress the secondary electron current that appears in the tetrode tube by the introduction of still another grid into the tube structure. This third grid is positioned between the plate and the screen grid, as indicated in Fig. 373. This fifth electrode, called the **suppressor grid,** is commonly connected to the cathode, as shown. Because the suppressor grid is negative with respect to the plate, it repels the secondary electrons so that they are recaptured by the plate. This feature makes it possible to utilize a wide swing in the plate potential

and thus obtain a relatively high power output from the tube. Further, in r-f amplification, the suppressor grid makes it possible, due to additional shielding of the plate, to secure high-voltage amplification even when using comparatively low plate potentials. Indeed, the amplification may reach a value of 1,500. Because the inverse electronic current

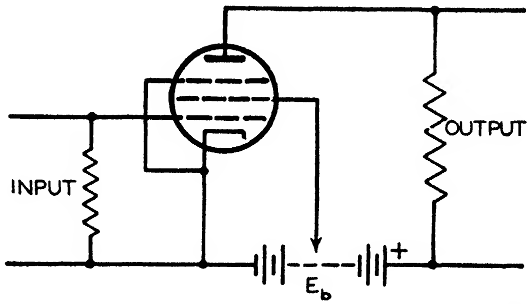


FIG. 373.—Functioning of a five-electrode tube, the pentode. Note that the suppressor grid is connected to the cathode.

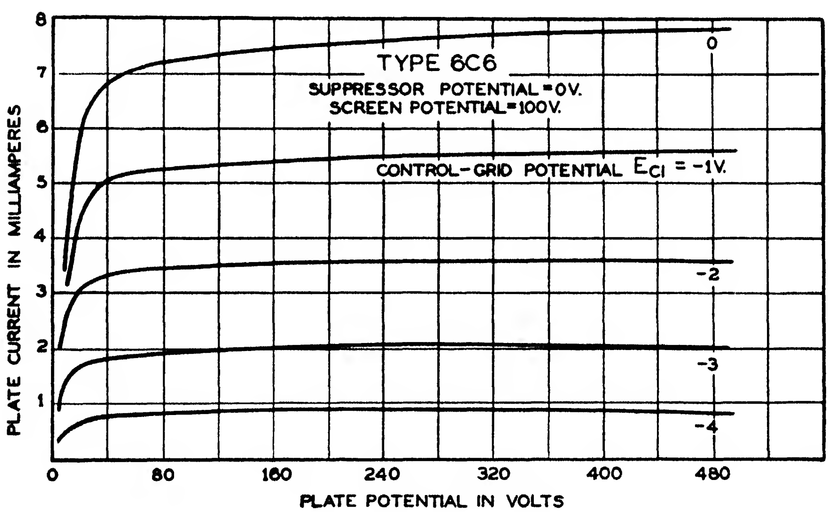


FIG. 374.—Operating characteristics of a pentode. Compare these curves with those for the tetrode, as shown in Fig. 372.

is eliminated, the characteristic curve for the **pentode** (as this tube is called) does not show the dip that appears in the curve for the tetrode. A family of pentode curves for a 6C6 tube is shown in Fig. 374. In the pentode the screen grid is operated at about half the potential of the anode, as in the tetrode. Pentodes are used chiefly in r-f amplifying circuits, though they are sometimes used as audio-power amplifiers.

It is possible to utilize a pentode as a triode by connecting the suppressor grid and the screen grid to the plate. As a triode, such a tube would have a μ value of the order of 20.

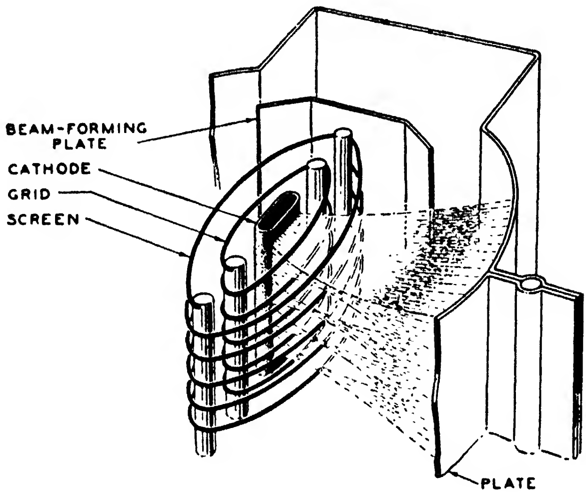


Fig. 375.—Beam-power tube.

An electron tube, known as the **beam-power tube,** has been developed that brings about the suppression of secondary emission effects without the use of a third grid, as in the pentode. Physically this tube is a tetrode; but it functions as a pentode, and has certain advantages over the conventional pentode. The suppression of secondary electrons is accomplished by so arranging the electrodes that rather definite electron beams are formed, as indicated by the diagram appearing as Fig. 375. (Beams are actually formed on both sides of the structure, though not shown in the diagrammatic sketch.) It will be seen that there is a region where a concentration of moving electrons exists. Such a concentration amounts to a **space charge** and a consequent **low-potential region.** The secondary electrons emitted by the plate will accordingly be repelled and eventually recaptured by the anode. Thus the space-charge region performs the function of a suppressor grid. The screen grid is so positioned that it is in the electron shadow of the control grid, with the result that the screen-grid current is low. Since no physical suppressor grid is present to obstruct the flow of electrons, all electrons

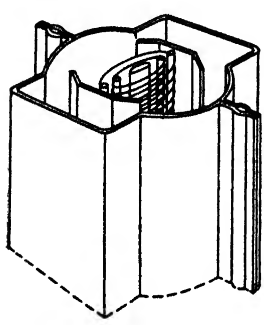


Fig. 376.—Structure of the beam-power tube.

will reach the plate, thus augmenting the tube output. The tube has two special plates that are connected to the cathode, their function being to assist in the confining of the beam to rather definite regions. A conventional diagram showing the spatial relations of the several tube elements is sketched in Fig. 376. Beam-power tubes will deliver a greater power output than any other tube of corresponding size; they are operated at plate potentials ranging from 250 to 400 volts.

As a further example of thermionic tubes, mention may be made of the **electron multiplier.** We have seen that, in the case of certain tubes, secondary emission proves to be a troublesome feature. It has been found possible, however, to **utilize** secondary emission advantageously. In Fig. 377, P is a photoelectric surface; P_1, P_2, and P_3 are electrodes

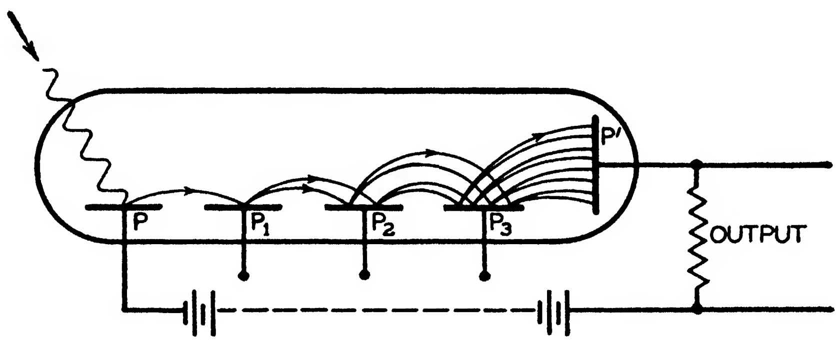


FIG. 377.—Electron multiplier. (*Courtesy of Proceedings of the Institute of Radio Engineers.*)

whose surfaces readily emit secondary electrons; and P' is the conventional anode. Each successive emitting surface is maintained at a higher potential than the preceding one. Luminous energy incident on P liberates, say, one electron. This will leave the emitting surface normally. An externally applied magnetic field (not shown) causes this electron to travel in a curved path until it strikes P_1. For simplicity's sake let us suppose that this impinging electron liberates two secondary electrons. These two in turn are magnetically deflected so that they strike P_2. At this electrode each will in turn liberate two other electrons, and so on, until, after leaving the last emitting surface, the beam of electrons impinges on the anode P' and thus becomes the output current. In the simple case cited, eight electrons will be delivered to the output electrode. It is not difficult to arrange conditions so that as many as five secondaries are liberated for each impinging entity, and as many as ten such cathodes are feasible. Under those circumstances, the total multiplication factor would be 5^{10}. A tube giving a multiplication of

that order would deliver sufficient energy at the output to operate a moderately large loud-speaker without additional amplification.

Such a plan of multiplication is, of course, applicable to an assembly in which the primary emission is thermionic instead of photoelectric. In such a case a control grid would be located near the cathode, and the organization as a whole would function as an extremely compact multi-stage amplifier. Electron-multiplier tubes are in practical operation.

Perhaps the most unique electron tube yet designed was developed under the direction of Dr. V. K. Zworykin. It consists of an electron gun and a light-sensitive mosaic, and is called the **iconoscope.** Figure

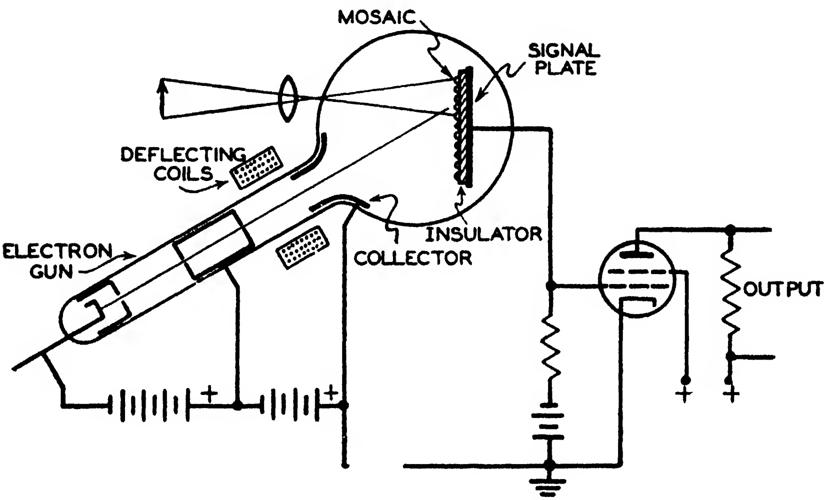


Fig. 378.—Diagram showing the essential components of the iconoscope and associated circuits. (*From a paper by V. K. Zworykin, "The Iconoscope," courtesy of Proceedings of the Institute of Radio Engineers.*)

378 is a diagrammatic representation of the design and method of operation of this modern "electric eye" which is now being used in the rapidly expanding field of television.

The most important feature of the iconoscope is the light-sensitive mosaic, which, in certain respects, bears a striking resemblance to the retina of the eye. The sensitive surface consists of a large number of tiny caesium-on-silver photoelectric areas. These minute light-sensitive, or picture, elements are electrically insulated from one another by being deposited on a common surface such as a sheet of mica. The mica sheet is backed by a thin metal film. Each tiny light-sensitive area thus becomes one element of a condenser; the metal backing sheet forms the other condenser electrode.

When light, in the form of an optical image, is incident on the mosaic

plate, each light-sensitive element emits electrons in proportion to the intensity of the illumination at that point. These photoelectrons are collected by an anode and pass around the circuit, as in the ordinary photoelectric cell. Since certain of the photoelements lose electrons, these elemental condensers manifest a positive charge on the image side of the receiving plate. The magnitude of the charge on any given condenser element will be proportional to the intensity at that point and to the length of time the image is allowed to fall on the surface. In the ordinary photoelectric tube the liberated electrons are replaced through a

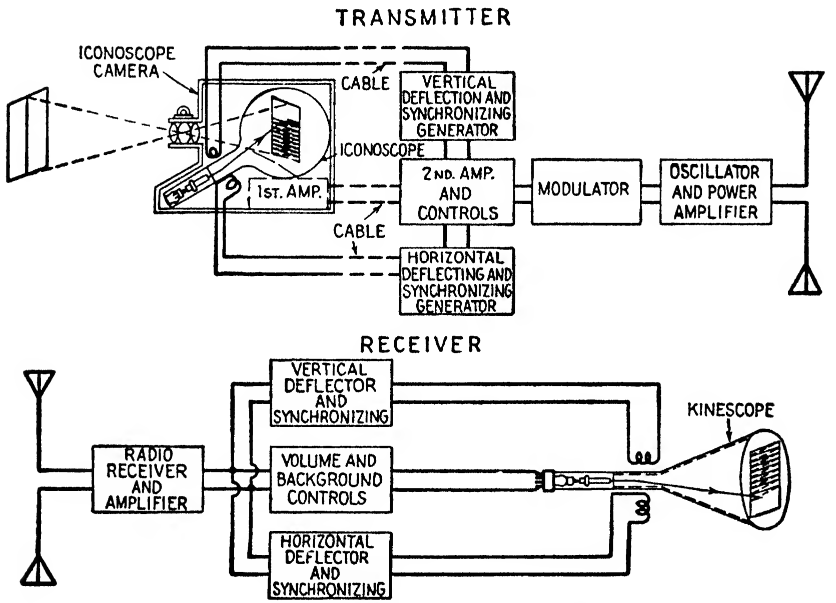


FIG. 379.—Block diagram showing the essential elements of a television transmitting and receiving system. (*From a paper by V. K. Zworykin, "The Iconoscope," courtesy of Proceedings of the Institute of Radio Engineers.*)

connection to a local d-c supply. In the iconoscope the replacement is brought about by means of a suitably focused cathode beam that is caused to scan the entire sensitive surface. This scanning process consists in a succession of horizontal sweeps (441, as of this date) from top to bottom of the image plate, the entire plate being covered in $\frac{1}{30}$ sec. As the cathode beam touches a positively charged light-sensitive element the electron deficit is replaced. This has the effect of discharging that particular elemental condenser with the result that a potential pulse will be conveyed to the grid of an associated amplifier tube, and thence to a transmission circuit. It will thus be seen that, as the cathode sweeps

over the image area, a succession of potential pulses will be delivered to the amplifier, and that a complete set of such impulses will occur 30 times per second. The output of the video amplifier is caused to modulate a carrier wave that, in turn, serves to control the intensity of an electron beam of a cathode-ray tube as it sweeps over the receiving screen, thereby reproducing the original image. Thus we have a television circuit. A block diagram of a television transmitting and receiving layout is reproduced in Fig. 379. The reader will find a series of eight technical papers in the August, 1945, number of the *Proceedings IRE* devoted to the theory and functioning of television equipment.

There are various other thermionic tubes that have important specialized uses, the theory and design of which will be found treated in books devoted entirely to electronic subjects.

PROBLEMS

1. Using the graphs shown in Fig. 350, determine the value of the amplification factor when a 6C5 triode is operated at the three plate potentials shown, and when the plate current is 8 ma. Also determine its value when the plate current is 4 ma.

2. Using the curve given in Fig. 351, determine the dynamic plate resistance for the 6C5 tube when the grid bias is −8 volts and the plate current is 2, 4, and 8 ma, respectively.

3. Again referring to Fig. 350, determine the transconductance of the 6C5 when the plate potential is 250 volts and the plate current 2 ma. Also indicate the transconductance when the plate current is 8 ma.

4. Using the values for the amplification factor and the dynamic plate resistance found in Probs. 1 and 2 above, compute the value of the transconductance from those factors, and compare with the result found in Prob. 3.

5. Compare the values that you found for the tube coefficients in Probs. 1, 2, and 3 with the values as read from the chart shown as Fig. 353.

6. If the plate resistance of the 6C5 is 10,000 ohms and the amplification factor is 20, what voltage amplification may be expected if the tube operates into a load resistance of 10,000 ohms?

7. Suppose that a signal whose rms value is 35 volts is impressed between the grid and the cathode of a triode whose amplification factor is 4.2, and which operates into a load resistance of 2,500 ohms. What a-c power will the tube deliver?

CHAPTER XXX

ELECTROMAGNETIC WAVES AND SOME APPLICATIONS

242. Equations of Electric and Magnetic Fields Referred to Rectangular Coordinates. Most students who pursue the course in electricity and magnetism, for which this text is intended to serve as a basis, will probably not have the mathematical background which will enable them to follow in detail the mathematical procedure involved in a consideration of Maxwell's theory of electromagnetic phenomena. It is possible, however, for any reader who has carefully read the preceding chapters to follow an outline of this important theory.

Faraday was not skilled in the processes of mathematical analysis; he did however possess what amounted to a prophetic insight into the processes of nature. Faraday could not conceive of "action at a distance"; to his way of thinking, some form of medium was involved in all electrical and magnetic effects that make their appearance at a distance from the primary cause of the phenomenon. Maxwell gave mathematical form to Faraday's concepts, and thereby evolved a theory that pointed the way to significant conclusions, and ultimately to important practical results. Let us glance at the analytical technique whereby Maxwell attained these ends.

Let us suppose that we have a region in which an electrostatic or magnetic field obtains. Consider an elemental volume in such a field, as sketched in Fig. 380. Let $\mathcal{E}$ be the electric intensity (es field strength) at the point p. By applying Gauss's theorem (Sec. 9) to such an elemental electrostatic field, it may be shown that

$$\frac{d\mathcal{E}_x}{dx} + \frac{d\mathcal{E}_y}{dy} + \frac{d\mathcal{E}_z}{dz} = \frac{4\pi\rho}{K}, \tag{286a}$$

where such terms as $d\mathcal{E}_x/dx$ signify the **rate of change of $\mathcal{E}$ in a direction parallel to the specified axis**; ρ is the **volume density** of the charge within the volume $dx\ dy\ dz$; and K is the dielectric constant.

Likewise, for a corresponding elemental magnetic field,

$$\frac{dH_x}{dx} + \frac{dH_y}{dy} + \frac{dH_z}{dz} = \frac{4\pi\mu}{\delta}, \tag{286b}$$

where H indicates magnetic-field strength, δ the **volume density** of magnetism, and μ the permeability. Obviously, if ρ and δ are zero, the sum of the space rates in the above two expressions becomes equal to zero.

Next let us consider the situation from the standpoint of any current that may exist in our hypothetical field. We have shown (Sec. 123), in

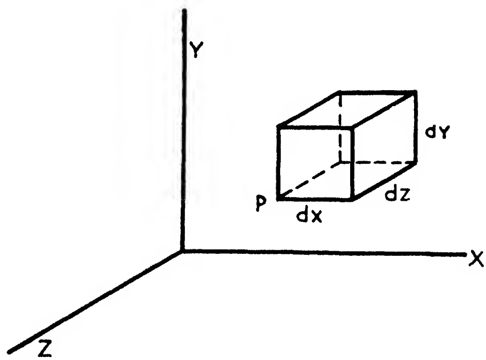


FIG. 380.—Elemental volume of an electromagnetic field.

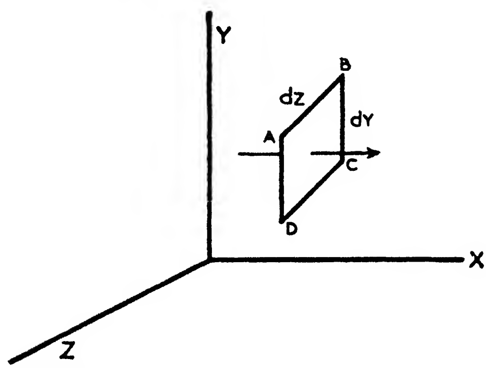


FIG. 381.—Current in an electromagnetic field.

general, that the work done in carrying unit pole around a closed path along which a current is flowing is numerically equal to $4\pi I$. Assume a path $ABCD$ (Fig. 381) about a current whose density per unit area is I. On this basis it may be deduced that

$$\frac{dH_z}{dy} - \frac{dH_y}{dz} = 4\pi I_x \tag{287a}$$

$$\frac{dH_x}{dz} - \frac{dH_z}{dx} = 4\pi I_y \tag{287b}$$

$$\frac{dH_y}{dx} - \frac{dH_x}{dy} = 4\pi I_z, \tag{287c}$$

where the I terms signify the current density in the directions indicated; the H terms the magnetic-field intensity in directions parallel to the three axes; and the differential coefficients the space rate at which these components vary.

Still another set of relations may be obtained by making use of two facts: (1) that the magnitude of an induced emf is given by $-d\phi/dt$; and (2) that emf can be expressed in terms of work, which in turn involves electric flux and distance. The application of these principles leads to

$$\frac{d\mathcal{E}_z}{dy} - \frac{d\mathcal{E}_y}{dz} = -\mu \frac{dH_x}{dt} \qquad (288a)$$

$$\frac{d\mathcal{E}_x}{dz} - \frac{d\mathcal{E}_z}{dy} = -\mu \frac{dH_y}{dt} \qquad (288b)$$

$$\frac{d\mathcal{E}_y}{dx} - \frac{d\mathcal{E}_x}{dy} = -\mu \frac{dH_z}{dt}. \qquad (288c)$$

It is evident that we now have three sets of field equations based on well-known fundamental concepts. It is possible to work out a fourth set of relations which, when combined with those already developed, will lead to several important deductions.

It is at this point that Maxwell made his most significant contribution. Before proceeding further it would be well for the student to reread Sec. 8. In that discussion it was pointed out that, if and when the charge on the bounding surfaces changes in value, the electrons of a dielectric undergo a displacement within the atoms. This electronic movement constitutes what might be spoken of as a form of displacement current **in the dielectric**; it results from the polarizing process. But this is not the whole story. Maxwell's conception of the case was to the effect that **all electric circuits are closed.** Even though the two plates of a condenser are separated by free space, it was Maxwell's thought that the ether undergoes a distortion due to the existence of the electrostatic field. He further held that any change in this condition constitutes, in effect, a current, and that this "space" current would give rise to the usual magnetic effects. According to this reasoning, then, one might expect to observe a changing magnetic field in the region between the plates of a condenser that is charging or discharging; and this even though a **vacuum** exists between the condenser electrodes. As used by Maxwell, the term **displacement current** was applied to the hypothetical space current just referred to. It is possible to derive an expression for such a space current.

To carry out such a derivation let it be assumed that the area of each plate of our condenser is A, and that the plates are very near together. Further, assume that at any instant each plate bears a charge Q. Under these conditions it would follow that the total number of lines of force

between the plates would be given by

$$\psi = 4\pi Q, \tag{i}$$

and the number of lines per unit area would be

$$\frac{\psi}{A} = \frac{4\pi Q}{A} = D. \tag{ii}$$

By definition, the flux density D is numerically equal to the field strength $\mathcal{E}$. We therefore have

$$D = \mathcal{E} = \frac{4\pi Q}{A}. \tag{iii}$$

If now we find the time rate of change of both $\mathcal{E}$ and Q, we have

$$\frac{d\mathcal{E}}{dt} = 4\pi\left(\frac{1}{A}\right)\left(\frac{dQ}{dt}\right). \tag{iv}$$

If the space between the plates contains a dielectric, it follows from Eqs. (5) and (6) that

$$\frac{d\mathcal{E}}{dt} = \frac{4\pi}{K}\left(\frac{1}{A}\right)\left(\frac{dQ}{dt}\right). \tag{v}$$

But

$$\frac{dQ}{dt} = i, \tag{vi}$$

which is the current passing to the condenser as the charging process takes place. From (iv) it may be set down that

$$\frac{1}{A}\left(\frac{dQ}{dt}\right) = I, \tag{vii}$$

where I is the space-current density through the surface. From (v) we may therefore write that

$$\frac{d\mathcal{E}}{dt} = \frac{4\pi}{K} I. \tag{viii}$$

From (viii) it follows that

$$I = \frac{K}{4\pi}\left(\frac{d\mathcal{E}}{dt}\right). \tag{ix}$$

In free space K becomes unity.

The last equation is the mathematical embodiment of the central concept of Maxwell's electromagnetic theory. It means that, if **at any point in free space** the electric field $\mathcal{E}$ is changing with time, we may deduce from this that a hypothetical current exists whose density is I,

and that this space current has a direction determined by the direction of the changes in $\mathcal{E}$. Are the Maxwellian displacement currents realities? If so, do they produce magnetic effects? In other words, does the last equation above have any significance when we are dealing with free space, *i.e.*, when $K = 1$? In 1929 Van Cauwenberghe proved experimentally that a current, as defined by the relation

$$I = \frac{1}{4\pi}\left(\frac{d\mathcal{E}}{dt}\right),$$

actually may exist, and that such a current will develop a magnetic field. Further, as we shall see shortly, the existence of electromagnetic waves supplies convincing evidence that Maxwell's assumptions were correct.

In general $d\mathcal{E}/dt$, and hence I, may have any direction. However, we can set down their components that are parallel to our axes of reference, thus:

$$I_x = \frac{K}{4\pi}\left(\frac{d\mathcal{E}_x}{dt}\right) \qquad I_y = \frac{K}{4\pi}\left(\frac{d\mathcal{E}_y}{dt}\right) \qquad \text{and} \qquad I_z = \frac{K}{4\pi}\left(\frac{d\mathcal{E}_z}{dt}\right).$$

By the aid of the above relations, Eqs. 289*a*, *b*, and *c* may be rewritten in the form

$$\frac{dH_z}{dy} - \frac{dH_y}{dz} = K\,\frac{d\mathcal{E}_x}{dt} \tag{289a}$$

$$\frac{dH_x}{dz} - \frac{dH_z}{dx} = K\,\frac{d\mathcal{E}_y}{dt} \tag{289b}$$

$$\frac{dH_y}{dx} - \frac{dH_x}{dy} = K\,\frac{d\mathcal{E}_z}{dt}. \tag{289c}$$

These relations constitute our fourth set of equations.

The two foregoing sets of equations (288 and 289) should be carefully examined and compared. It is to be noted that all six of these equations involve **rates of change** of electric and magnetic flux, but the **dielectric constant** K appears in one set while the factor of permeability enters into the other group. Since both sets of equations contain $\mathcal{E}$ and H terms it is possible, by the aid of the (286) group, to combine these relations into one set of expressions—a group of equations which led Maxwell to make an exceedingly important deduction.

If, for instance, one begins with Eq. (289*a*), and proceeds as suggested above, the result is

$$\frac{d^2\mathcal{E}_x}{dx^2} + \frac{d^2\mathcal{E}_x}{dy^2} + \frac{d^2\mathcal{E}_x}{dz^2} = K\mu\,\frac{d^2\mathcal{E}_x}{dt^2}, \tag{290}$$

Treated in like manner, the other (289) equations, will yield corresponding expressions. Now it can be shown that the general equation

for **wave motion** has the above form **except for the term $K\mu$.** In the wave equation the corresponding factor is $1/v^2$, where v is the velocity of the wave disturbance. If, then, $v = 1/\sqrt{K\mu}$, Eq. (290) would represent a **wave motion.** It is known that the term $1/\sqrt{K\mu}$ is numerically equal to the velocity of light. It thus becomes evident that concomitantly changing electric and magnetic fields constitute **a wave disturbance in space,** and that the wave velocity is the **same as the velocity of visible radiation.** Thus it becomes evident that electric and magnetic forces do not **instantly** produce an effect at a distance but are propagated through space with a finite velocity, *i.e.*, with the velocity of light.

The significance and the utility of Maxwell's equations will become apparent if we apply them to the case of a plane wave in which the intensity at any one instant is the same over the entire wave front. For convenience, let us assume that the wave front is in the yz plane (Fig. 381), and that the wave is traveling in the positive direction of x. Under these assumptions, the differential coefficients with respect to y and z will be zero. On that basis the 288 equations become

$$-\mu \frac{dH_x}{dt} = 0 \qquad -\mu \frac{dH_y}{dt} = \frac{d\mathcal{E}_x}{dx} \qquad \text{and} \qquad -\mu \frac{dH_z}{dt} = \frac{d\mathcal{E}_y}{dx}.$$

From the first relation of the above group it follows that H_x is zero, which means that there is no magnetic vector in the x direction.

By proceeding in a similar manner with our 289-equations we find that

$$K \frac{d\mathcal{E}_x}{dt} = 0 \qquad K \frac{d\mathcal{E}_y}{dt} = -\frac{dH_z}{dx} \qquad \text{and} \qquad K \frac{d\mathcal{E}_z}{dt} = \frac{dH_y}{dx}.$$

From the first relation above we see that $\mathcal{E}_x = 0$; hence there is no electric component in the x direction. From the foregoing we may conclude, therefore, that both the magnetic and the electric vectors are **in the plane of the wave.** It remains to be seen what spatial relation these two vectors bear to each other.

Let us assume that the electric vector $\mathcal{E}$ is parallel to the Y-axis. On that assumption, $\mathcal{E}_y = 0$. From the second equation of the last group it is to be seen that when $\mathcal{E}_y = 0$, H_y is also zero. There remain, then, only $\mathcal{E}_z$ and H_y, which means that the electric vector is parallel to the Z-axis and that the magnetic vector is parallel to the Y-axis. These two vectors are, therefore, at right angles to each other, and to the direction of propagation. From the last equation of the last group it is evident that the electric and magnetic vectors (electric and magnetic field strengths) are **interdependent;** a variation in the magnitude of one gives rise to a change in the other.

In concluding our consideration of Maxwell's equations and their relation to the propagation of electromagnetic waves, the energy relations should be noted. In Sec. 16 it was established that the energy content per cubic centimeter of an electric field is given by the expression $K\mathcal{E}^2/8\pi$, and in Sec. 32 we saw that the corresponding energy in a magnetic field is $\mu B^2/8\pi$. It may be shown that $\mu H^2 = K\mathcal{E}^2$, from which it follows that, at any instant, the energy in an electromagnetic wave is equally divided between the two fields. The **total energy per unit volume** can therefore be expressed in several ways, thus,

$$\frac{\mu H^2}{8\pi} + \frac{K\mathcal{E}^2}{8\pi} = \frac{\mu H^2}{4\pi} = \frac{K\mathcal{E}^2}{4\pi} = \frac{\sqrt{\mu K}}{4\pi} H\mathcal{E}.$$

The above expressions indicate that, at any instant, the energy per unit volume of the medium through which the wave is moving is proportional to H^2, to $\mathcal{E}^2$, or to the product of the electric and magnetic vectors.

The introduction of the electromagnetic theory of wave propagation was an epoch-making step in scientific history; and, as we shall presently see, far-reaching results have followed. It should, however, be noted in passing that Maxwell's theory **does not** account for several important phenomena that are encountered in the field of optics. One such instance lies in the fact that the theory does not account for the **emission or absorption** of radiant energy. But even though the theory fails in certain particulars it has served a useful purpose in pointing the way to investigations that have resulted in discoveries that have profoundly influenced the history of mankind.

For a more complete discussion of this subject the reader is referred to a monumental work by the late Professor J. A. Fleming, "Electric Wave Telegraphy and Telephony," Chap. V. For a treatment based on the vector method, see "Principles of Electricity and Magnetism" by Professor G. P. Harnwell.

243. Discovery of Electromagnetic Waves. Maxwell introduced his electromagnetic theory of wave propagation in 1865. For many years the only direct evidence in support of the theory was the experimental confirmation of the relation between the known velocity of light and the factors μ and K.

In 1880, Professor Heinrich Rudolph Hertz,[1] a German physicist, made the highly important announcement that he had been able to produce, and to measure, electromagnetic waves.

[1] Hertz was born at Hamburg on Feb. 22, 1857, and died at the early age of 37. He graduated from the University of Berlin where he was a favorite student of Von Helmholtz. It was while he was a professor of physics at the Technical College of Karlsruhe that he carried out his epoch-making research.

Hertz made use of the oscillating discharge of a capacitance as a generator of such waves. It has already been shown that the discharge of an electrified capacitance through a circuit containing an inductance and a resistance of small magnitude gives rise to electrical oscillations whose period is given by the expression $2\pi\sqrt{LC}$. The setup employed by Hertz consisted of an induction coil to the terminals of which were connected capacitance areas A and A', as sketched in Fig. 382. When a discharge took place at the spark gap S, Hertz found that he could detect electrical effects at a distance of several meters. His detecting device consisted of a single loop of wire that was all but closed by a micrometer spark gap. When his oscillator was in operation he found that, under certain conditions, tiny sparks would appear in the gap of his receptor loop, thus indicating that energy was being transmitted **through the intervening space** between the oscillator and the detecting device. The constants of the circuits employed by Hertz were of such a magnitude that the oscillations had a frequency of the order of 5×10^9. This frequency corresponded to a wave length in the region of 60 cm, though in certain of his experiments he used waves considerably shorter than that. In fact, all of Hertz's original work was done with waves that we now refer to as "high-frequency radiation." Because of the resistance of the spark gap that formed a part of Hertz's generator, the oscillations were highly damped; each spark that occurred gave rise to oscillations of the character shown in Fig. 383.

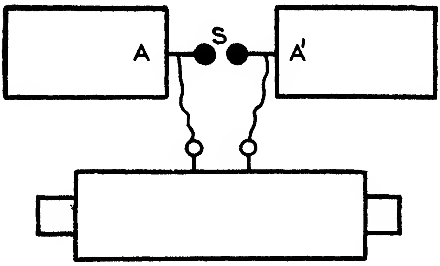


FIG. 382.—Assembly used by Hertz in studying electromagnetic waves.

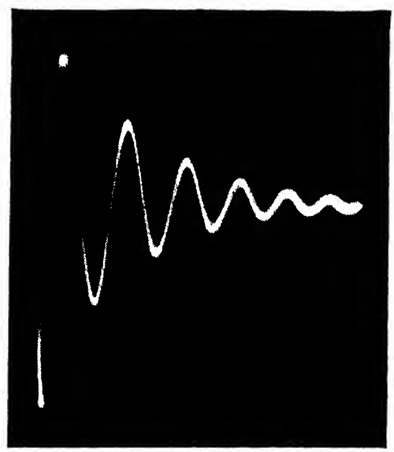


FIG. 383.—Dampened electrical oscillations. *(From "A Textbook of Physics" by C. A. Culver, courtesy of The Macmillan Company.)*

Notwithstanding the limitations under which he worked, Hertz experimentally established the following facts in addition to proving the existence of the form of radiation predicted by Maxwell:

1. Electromagnetic waves may be reflected (the basis of present-day radar).
2. Rectilinear propagation occurs.
3. Such waves can be refracted by nonmetallic mediums.
4. They can be polarized.
5. Diffraction phenomena may be produced.

Thus it will be seen that, though much longer than light waves, electromagnetic waves obey all of the laws that obtain in the case of light.

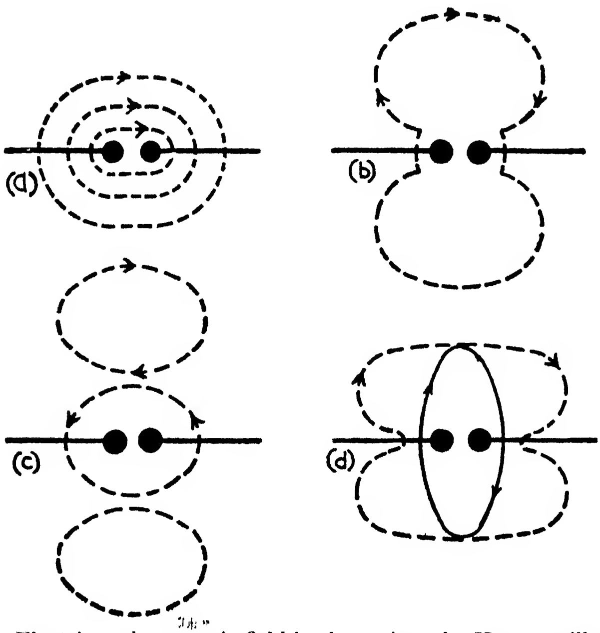


FIG. 384.—Electric and magnetic field in the region of a Hertz oscillator. Broken curves indicate lines of electric force; heavy trace shows a magnetic line of force.

While we now make use of undamped waves in radio communication, we frequently utilize a radiating system that is essentially the same as that employed by Hertz. For this reason, and also because general principles are involved, it will be worth while to glance at the electric and magnetic field conditions as they exist in the region of a Hertz oscillator.

Referring to Fig. 384, as the capacitance areas acquire a charge, the electric field may be represented as in (*a*). If the potential difference is increased until a spark passes between the terminals, there will be a sudden movement of charges along the conductors, that will, in

effect, bend the lines of force, as indicated in (*b*). Since the discharge is oscillatory, the electrons will move to and fro along the conductor with the result that loops of strain will be, as it were, "snapped off" as shown in (*c*), and a new set of lines of force will form. It is to be noted that the direction of the electric strain in the new loop is opposite to that of the original loop. (Why?) These alternations in the electric field travel outward with a finite velocity equal to $1/\sqrt{K\mu}$. Remembering now that any change in the value of $\mathcal{E}$ will give rise to a magnetic field, we see that magnetic lines of force will appear as indicated in (*d*). Thus we have concomitant electric and magnetic fields, both alternating in character, and each at right angles to the other. Actually today we use conducting rods a few centimeters in length as oscillating elements, but we generate the electric oscillations by means of electron tubes rather than by means of a spark gap, and thus develop continuous undamped electromagnetic waves.

244. Marconi's Contribution. Following the disclosures made by Hertz, a number of investigators made attempts to utilize these so-called **Hertzian waves** for the purpose of transmitting intelligence through space, but without marked success. Improved detecting devices, however, were developed by Lodge and Branly. The record is not clear as to when Marconi first began the researches which eventually led to a practical system of space telegraphy. We do know that his experiments were in progress in 1895, and that his first trials were made at or near Bologna, Italy. At first Marconi utilized the Hertzian type of oscillator, consisting of two capacitive areas and a corresponding receiving system. He, however, made use of a more sensitive detecting device than did Hertz, viz., the Branly-Lodge coherer. It occurred to Marconi that one might utilize the **earth** as one of the two capacitive areas of a Hertz oscillator system. **Herein lay Marconi's great contribution to "wireless telegraphy."** Eventually he used a simple wire as one member of what amounted to a condenser, the earth forming the other electrode. A similar arrangement was utilized as a receiving system, as shown in Fig. 385.

In 1896, Marconi went to England and applied for his first patent. In the following year he demonstrated transmission over water up to 10 miles or more, and in 1899, he established communication across the English Channel, a distance of 32 miles (50 km). From this point on, the story of space communication unfolds rapidly, and the end has not yet been reached.

The reason for giving the foregoing brief review of Marconi's contribution to this all-important means of communication has been to point out that various able investigators, over a period of some fifteen

years, had striven to make a practical application of the discoveries of Hertz but largely had failed. Marconi, a young man in his early twenties, by the introduction of the earth connection, took the Hertz apparatus out of the laboratory and gave it world-wide application.

In passing, however, it should be recorded that Sir Oliver Lodge introduced what he called **syntonized** space telegraphy Lodge's invention was made in the early days of Marconi's transmission experiments and contributed materially to the success of those early tests. Lodge's contribution consisted of circuit arrangements whereby the radiating and receiving systems could be brought into electrical resonance (Sec. 164).

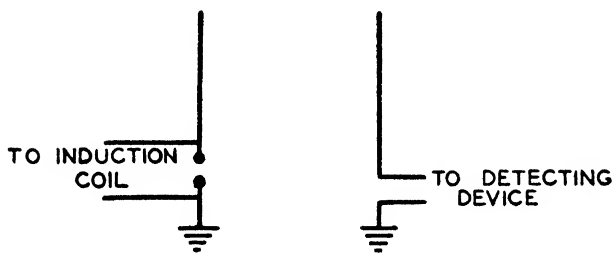


FIG. 385.—Type of radiating and receiving systems used by Marconi in his early experiments.

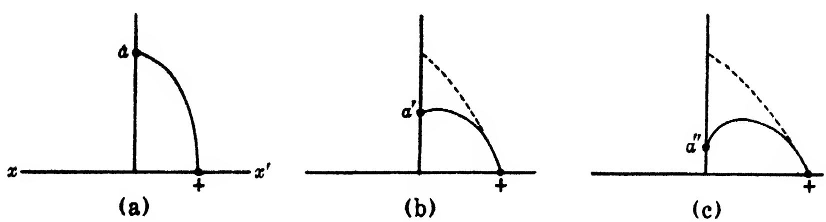


FIG. 386.—Process by which radiation takes place from a Marconi antenna.

245. Radiation from, and Absorption by, an Antenna. Let it be assumed that there is available some means whereby the free electrons in a simple vertical wire, connected to earth at its lower end, may be caused to oscillate in that conductor. Further, let us consider what happens when a single electron moves rapidly from the position a (Fig. 386a) toward the earth end. Electrostatic lines will exist between the electron and corresponding positive charges on the conducting plane, one such line being shown. If the electron moves to the position a' as shown in (b) the line of electric flux will be bent as indicated. Still further displacement toward the plane xx' is shown in (c). The distortion of the ether caused by the electronic motion will spread outward from the source with the velocity of light. Since we have assumed the to-and-fro movement of the charge to be of a high frequency, **all of the**

energy resident in the electrostatic field cannot return to the source before the electron is again moving through any given point in the same

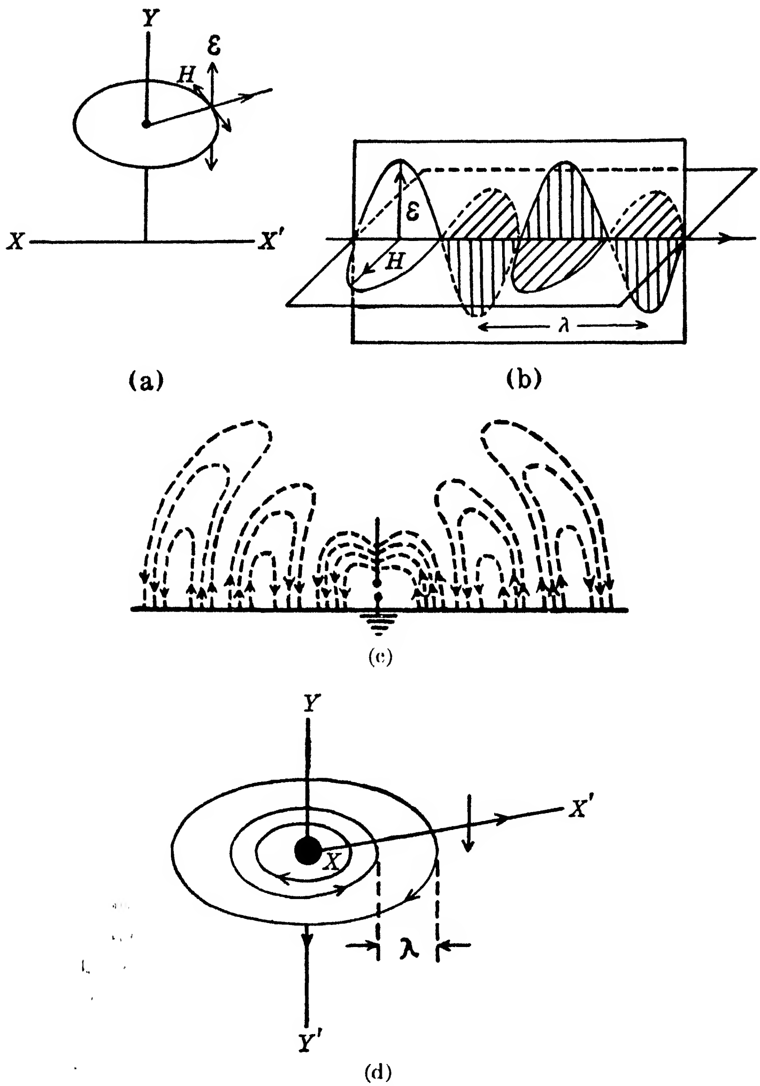


Fig. 387.—Showing the relation of the electric and magnetic vectors as radiation takes place.

direction. Hence a part of the energy given up to the field never returns to the source, but continues to move outward as a wave disturbance. At low (commercial) frequencies most of the energy in the

field has time to return to the source before reversal occurs; hence there is little radiation under these circumstances.

But, as we have previously seen, this is only a part of the story. As a line of electric flux sweeps through space it gives rise to a magnetic field. This generated magnetic field, in accordance with the principles already laid down, will be at right angles to the electric field and to the direction of its motion. The vector relations corresponding to both of these fields, at a given point in the field, is indicated in Fig. 387*a*, where $\mathcal{E}$ represents the electric vector and H the corresponding magnetic element. This created magnetic field is in time phase with the electric field but in space quadrature as shown in (*b*). As in the case of the electric field, the magnetic field cannot return all of its energy to the source between reversals of the electron, and so it continues to travel outward from the source, as does the concomitant electric pulse. Indeed, **these two fields are not to be thought of as separate and distinct disturbances but rather as two aspects of the same phenomenon.** If, then, we provide means whereby our electron will continue to oscillate —*i.e.*, if we continue to supply energy to the moving charge—it is evident that we may cause energy to be radiated into space in the form of electromagnetic waves. A vertical section of a detached group of the electrostatic pulses is shown in (*c*). If one were to view the plane XX' from the point Y the magnetic aspect of the field would consist of a series of expanding concentric circles, as indicated in (*d*). The manner in which continuous electronic oscillations are maintained in the radiating system will be indicated in Sec. 247. The frequencies currently employed range from 1,000 kc to more than 3,000 megacycles.

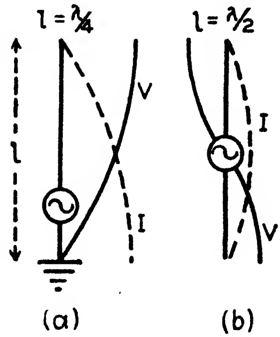


FIG. 388.—Current and potential relations (*a*) in the case of a quarter-wave Marconi antenna, and (*b*) of a Hertz radiator.

In dealing with the radiation from an oscillating system, it is necessary to take into account the form of the radiator. It is possible to set up standing electrical waves on conductors of finite length,[1] just as one may set up standing sound waves in an organ pipe. In the case of a radiator of the Marconi type (Fig. 388*a*), where the physical length of the conductor is equal to $\lambda/4$, there will be a potential node at the earth end and a current node at the upper end. On that basis, the funda-

[1] For an analytical discussion of the theory of electrical waves on conductors, see Chap. IV of Professor Fleming's treatise on "The Principles of Electric Wave Telegraphy and Telephony."

mental wave length is $4l$. If the radiator is excited at its mid-point, its simplest mode of resonant oscillation is as indicated in Fig. 388*b*; current nodes are to be found at each end, and a potential node at the center. This form of radiator is commonly referred to as a **Hertz** (or doublet-type) **antenna.** A Hertz antenna may be erected horizontally or vertically. The Marconi type is more suitable for long-wave radiation while the Hertz form is better adapted to short-wave operation. (Do

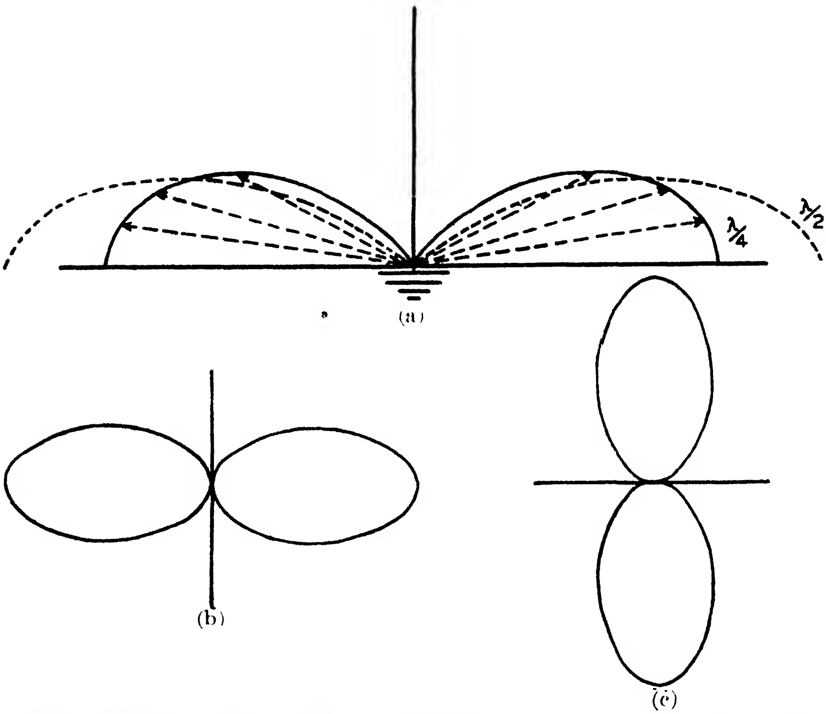


FIG. 389.—Radiation patterns of a grounded Marconi antenna and a Hertz radiator.

you see why?) Many broadcasting stations utilize a tower as a Marconi form of radiator.

The radiation characteristics of the Marconi and the Hertz types of antennas, as they are commonly used, are not the same. Figure 389*a* depicts approximately the radiation pattern in the case of an earthed vertical antenna whose physical length is $\lambda/4$. It is obvious that such an antenna emits low-angle radiation equally in all directions in a horizontal plane. (Radial distances from the base of the antenna are proportional to field intensity in that particular direction.) Analysis[1]

[1] For a complete discussion of the theory of radiation from a quarter-wave antenna, see p. 420, "Principles of Radio Engineering" by R. S. Glasgow.

shows that the field strength varies inversely as the distance from the base of the antenna (point of maximum current). Because the horizontal field is a maximum for a given power input, the half-wave antenna is widely used in connection with broadcasting installations.

The radiation patterns for the vertical and horizontal positions of the Hertz doublet are shown in Fig. 389*b* and 389*c*, respectively. Both positions are widely used for short-wave communication, the particular form depending upon the type of communication being carried on.

In dealing with emission phenomena, it is to be kept in mind that radiation from an antenna is, in general, polarized. **The polarization of an electromagnetic wave is taken to be the direction of the electric vector.** The horizontal radiation, for instance, from a Marconi-type antenna is **vertically polarized,** *i.e.*, its electric field is perpendicular to the surface of the earth. A wave is said to be **horizontally** polarized when its electrostatic component is **parallel to the earth's surface.** A horizontal Hertz-type radiator emits horizontally polarized waves. The magnetic vector will be at right angles to the plane of polarization.

Reference was made above to the matter of **field strength.** It has been previously pointed out (Sec. 242) that the energy radiated from an antenna is divided **equally** between the electric and the magnetic components. One might therefore express the field strength either in terms of the magnitude of the electric vector $\mathcal{E}$ or the magnetic vector H. If the first plan was followed, the field intensity would be expressed in volts per meter, as was done in the case of potential gradient (Sec. 43). If the latter method was used, the field strength might be expressed in terms of lines of magnetic flux per square meter, the area being tangent to the wave front at the point in question. It is the custom, however, to express this characteristic of the field by the first method, *i.e.*, in volts per meter. In practice, field strength, or field intensity, means the emf that will be induced in a conductor 1 m in length so placed that its axis is parallel to the direction of polarization. For example, if an emf of 5 microvolts is developed in a conductor when electromagnetic waves cut across it, the field strength is said to be 5 μv/m. In radio broadcasting a field strength of the order of 100 μv/m is considered necessary in order to give satisfactory radio reception.

When the alternating magnetic field of an electromagnetic wave moves across a conductor (a receiving antenna) an alternating potential difference is established between the terminals of the conductor. If the wire, together with any associated capacitance, is not in electrical resonance with the incident wave, the emf developed is extremely small. If, however, the receiving system is "tuned" to the same frequency as the traveling wave, a greater potential difference will be set up. The

alternating current that results from the alternating potential difference in the conductor may be amplified and subsequently converted into an audible or visual response.

Since, in general, an antenna system radiates more or less equally in all horizontal directions, it will be obvious that a "receiving" conductor absorbs only an exceedingly small amount of the total energy emitted by the radiating system; the over-all efficiency is extremely low. Efforts have been made to concentrate the energy emitted by a radiator into a beam, and thus increase the field strength in a given direction. This was first accomplished by Hertz. In certain of his original experiments he placed one of his doublets at the focal line of a cylindrical parabolic

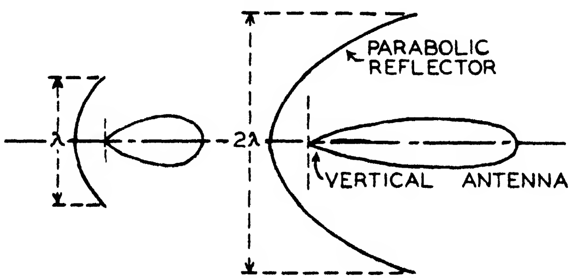


Fig. 390.—Radiation pattern produced by curved reflectors having different contours. (*From "Principles of Radio Engineering" by R. S. Glasgow, McGraw-Hill Book Company, Inc.*)

reflector, thus producing a parallel beam of radiant energy. The receiving conductor was also backed by a similar "mirror." It is interesting to note that we have in recent times reverted to this identical scheme in order to establish ultra h-f communication. This principle is utilized in connection with that form of radio communication designated as **radar** (Sec. 250). A diagrammatic representation of a reflector arrangement is shown in Fig. 390.

It is, however, not necessary to have a continuous metallic reflecting surface in order to bring about radiation in some particular direction. A relatively large percentage of the total energy emanating from a transmitting system may be concentrated into a beam by means of a system made up of simple conductors.

Imagine two simple vertical antennas located one-fourth wave length apart, and that we are looking at these aerials from above (Fig. 391). Further, assume that energy is supplied to antenna A and that A' is tuned to the same frequency as A, but receives no energy except by radiation from A. A' will be set into oscillation as a result of the incident energy received from A and will therefore reradiate a part of this energy. The energy reaching some distant receiving point to the right of A will

consist of two parts, energy coming directly from A and energy reradiated from A'. What phase relation will exist between these two wave trains? A wave in traversing the distance from A to A' will change (lag) in phase by 90°. In the process of generation in A' there will be another loss of 90°, and in reradiation a third change of 90° will take place. In addition to these three phase shifts, there will be a loss of 90° because of the time consumed by the wave in passing back from A' to A. We therefore have a total phase change of 360°, which means that the reradiated wave will be **in phase** with the wave which originally leaves A. The "reflected" waves will therefore serve to augment the energy being radiated directly from A and the field strength at any given receiving point p to the right of A will thus be increased.

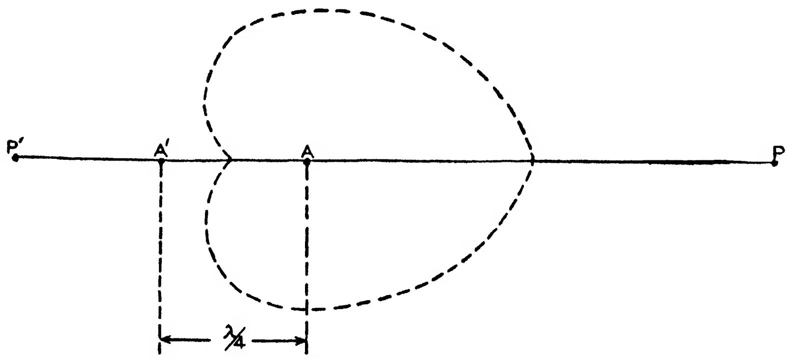


Fig. 391.—Concentration of radiation in a given direction by means of a "reflecting" antenna.

If we examine the phase relations behind the reflecting antenna wire, *i.e.*, to the left of A', we find that the energy radiated directly from A and that reradiated from A' arrive at any given point P' 180° out of phase, and hence the direct waves and the reradiated wave trains tend to nullify one another. This is due to the fact that the reradiation from A' is 270° out of phase with the wave emerging from A, and that the wave in passing from A to A' loses 90° in phase and will therefore differ by 180° from the wave leaving A'.

If the combined effects of a single antenna wire and one reflecting wire in directions other than those already discussed is examined and plotted, it will be found that the polar curve will take the well-known form shown in Fig. 391. If, however, an antenna "array," such as that sketched in Fig. 392, is employed, the field pattern is seen to be considerably narrower. It will be noted that the secondary radiators (reflector elements) are placed one-fourth of a wave length behind the primary radiators, and that a quarter wave length separates the adjacent elements. The length

of the individual conductors is equal to $\lambda/2$. The radiation represented by the small lobes on either side of the principal beam is, in general, inconsequential.

It will be apparent that if we concentrate the energy that would ordinarily be radiated throughout 360° within a beam whose limits are, say, 36°, the energy required to produce a given field strength at a certain receiving point would be but one-tenth that necessary without the reflecting system.

But the gain in efficiency does not end here. A reflecting system may also be utilized behind the receiving antenna. If the gain at the receiv-

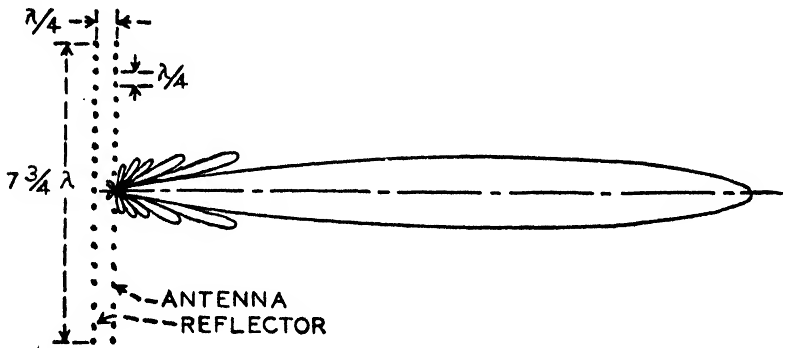


FIG. 392.—Radiation pattern due to a multiple-antenna array. The secondary maxima are negligible. (*From "Principles of Radio Engineering" by R. S. Glasgow, McGraw-Hill Book Company, Inc.*)

ing end is also on the basis of 1 to 10, the over-all efficiency of the system would be increased as the square, or one hundred fold. Experience has shown such an increase in efficiency to be entirely practicable.

In 1926 a two-way beam channel between England and Canada was put into commercial operation, utilizing a wave length of slightly over 26 m. This channel was arranged for duplex operation at a speed in excess of 100 words per minute. With this system a 20-kw transmitting plant will produce the same strength of signal at the receiving station as a 200-kw installation employing the older nondirectional radiating system. Since the Canada-England circuit was installed many other transoceanic beam channels have been established, some of them extending halfway around the world. The frequencies used on these long distance channels are, at present writing, in the 10 to 20 mc band, which corresponds to wave lengths ranging from 15 to 30 m. The directive arrays arranged for use in television are designed to operate at frequencies of the order of 300 mc ($\lambda = 10$ cm), and radar beams are operated at still higher frequencies. The advantages resulting from the use of the

higher frequencies are (1) greater radiation for a given power input (radiation varies inversely as the square of the frequency); (2) less electrical atmospheric disturbance in the higher frequency ranges; and (3) a greater number of available channels.

246. Propagation of Electromagnetic Waves. We have already discussed (Sec. 245) the principles involved in the radiation process. It remains to consider the *modus operandi* by which these waves reach distant points on or above the earth's surface.

The form of the moving wave front due to radiation from any given antenna system will depend, largely, on the electrical and mechanical characteristics of the radiator. For the sake of simplicity, let it be assumed that we have a hemispherical wave front. That portion of such a wave that is adjacent to the earth is commonly referred to as the **ground wave**; that part of the wave, whose radius vector makes a considerable angle with the earth's surface, say 45° ±, is spoken of as the **sky wave.** These waves are simply different sections of one wave front and **not** independent wave disturbances. The surface of the earth is a semiconductor; hence the ground wave undergoes attenuation. The magnitude of this attenuation is a function of the frequency; the higher the frequency, the greater the attenuation. This known fact gave rise to the use of relatively long waves (several thousand meters) in the early period of radio communication. But even this fact did not fully account for the success of Marconi's first transatlantic experiment. Why do even long waves follow the curvature of the earth?

Shortly after the First World War the United States government assigned the wave lengths of 200 m and below to the amateur radio operators; such frequencies were officially considered to be of no particular engiheering value. Because of the resulting congestion in the immediate vicinity of 200 m, amateurs began to experiment with the higher frequencies (shorter wave lengths) and, at least in some cases, were successful in transmitting messages over great distances when using quite low power. Soon research laboratories, following the path blazed by the youthful experimenters, discovered that the propagation of short waves did not follow the laws commonly applied to long-wave transmission. It was then found that reflection from the ionosphere (Sec. 203) was responsible for the phenomenal results attained when using the higher frequencies. The propagation process is shown diagrammatically in Fig. 393. When the angle made by the ray is not too great, the refraction in the ionosphere experienced by the sky wave results in the wave being bent back to the earth, as indicated in the sketch. The ground wave may be so attenuated in passing from the transmitter at T to an observer at R_1 that a signal could not be detected at that point; but a

refracted sky wave might deliver a readable signal at a much greater distance, say at R_2, or even R_3. While the return of the sky wave to earth is apparently due to a refraction process, it is convenient to think of it as a case of reflection, the "reflecting" plane being at a distance above the earth known as the **virtual** height. (See the dotted extension of the sky wave.) If the ground wave and the reflected wave were **both** to reach a given point **out of phase,** the interference effect might obliterate the signal. Since the height of the ionosphere is not a fixed quantity, this interference effect will vary, thus giving rise to **fading.** The wave band between 10 and 40 m has been found to be useful for sky-wave

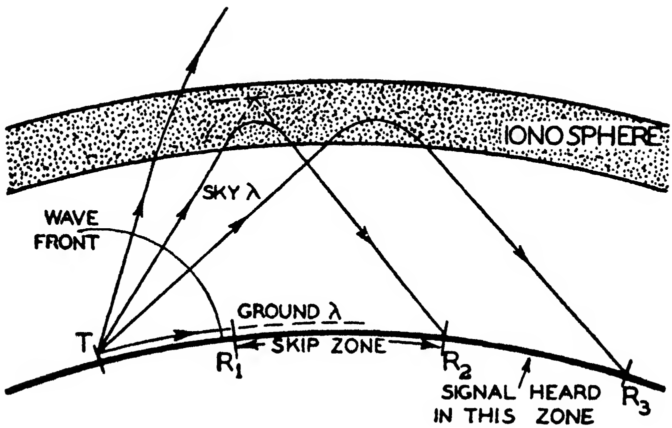


FIG. 393.—The ionosphere and its effect on radio transmission.

transmission. Waves below about ten meters are refracted by the ionosphere at an angle so great that the refracted beam misses the earth, and hence cannot be utilized for sky-wave transmission. Waves below approximately ten meters (30 mc) are accordingly useful only in those cases where "direct vision" is possible. Waves below 1 m (300 mc), and shorter, are useful for point-to-point communication, particularly since at those frequencies fading and atmospheric disturbances are largely absent.

247. Radio-transmitting Circuits. The several components which go to make up a radio-transmitting assembly have already been discussed. There are many forms of circuits utilized for the purpose of exciting a radiating system and modulating its output in conformity with an impressed signal. One simple transmitter layout is sketched diagrammatically in Fig. 394. The r-f circuit is shown in the upper part of the diagram, and the audio circuit in the lower. The first r-f tube A_1 functions as a crystal-controlled master oscillator of 5 to 50 watts output.

The second tube A_2, and its associated grid and plate circuits, is a

tetrode of about the same energy-handling capacity as the oscillator unit. A tetrode is used in order to avoid self-oscillation. The purpose of this stage is not to provide amplification, but to prevent any change in r-f load, due to modulation, from affecting the electrical stability of the master oscillator; hence it is called the "buffer" stage.

The third tube A_3, which is also a tetrode for the same reason as in the previous case, is designed to handle, perhaps, 50 watts. The anode of this tube is provided with a d-c plate supply B_6. The secondary of an audio power transformer T_4 is **in series** with this plate supply. This second r-f amplifier stage serves to drive the output r-f power stage

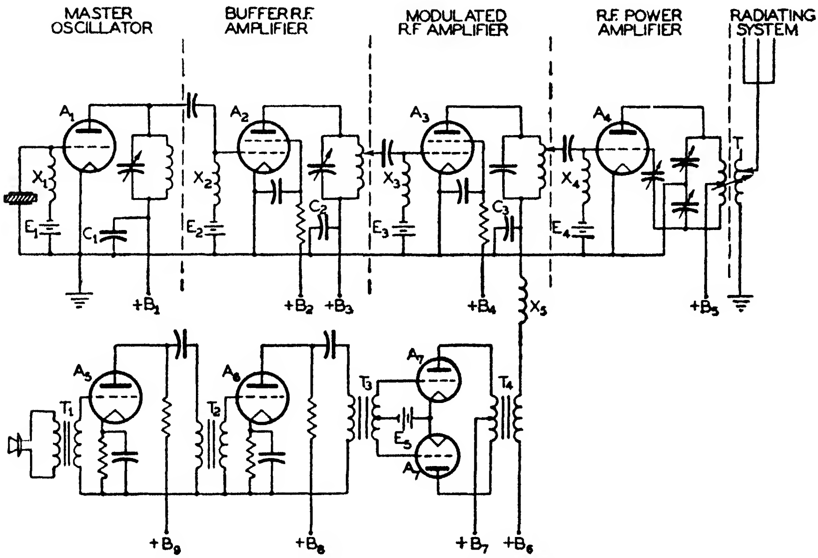


FIG. 394.—Schematic diagram of radiotelephone transmitter circuit employing amplitude modulation.

employing a tube A_4 which may deliver several hundred watts to the radiating system. If a still higher r-f output is required, one or more additional r-f amplifier stages would precede the output stage. Radio transmitters are now in use which will deliver in excess of 100 kw to the antenna system.

The audio-amplifying system, shown in the lower part of the diagram, consists of two or more voltage-amplifying stages, which in turn serve to drive a push-pull power stage. The output of this stage delivers a **signal-controlled voltage** to the plate circuit of the r-f power amplifier. Thus the potential applied to the anode of that r-f stage varies in conformity with the signal, and **modulation of the amplitude of the carrier wave is**

effected. If the signal consisted of vocal or musical sounds, the situation would be as depicted in Fig. 395.

Before leaving the subject of transmitting circuits, it will be noted that in the representative layout above discussed (Fig. 394) the last r-f (power)

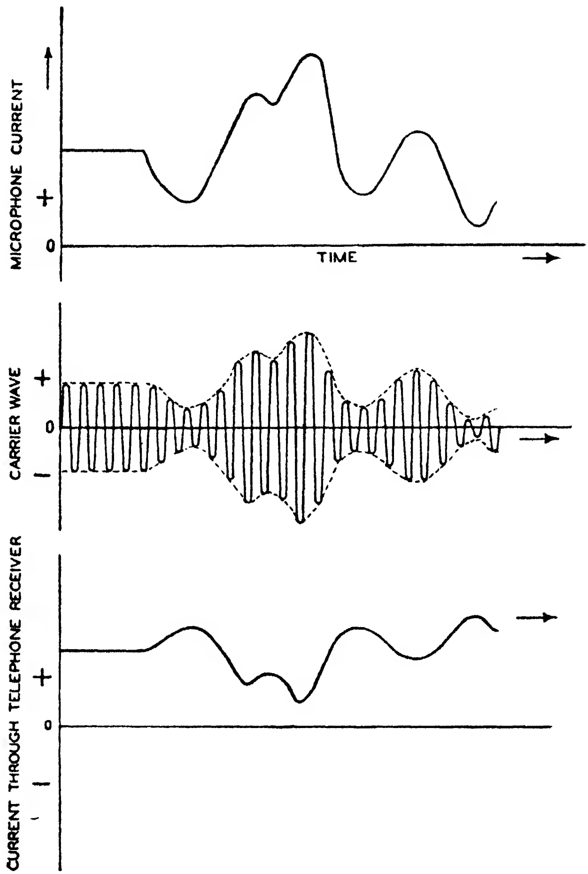


FIG. 395.—Showing the signal-wave form and the resulting modulated carrier wave. The last trace indicates the character of the output current after demodulation has taken place. This current is utilized for the purpose of operating telephone receivers or a loud speaker.

stage involves the use of a triode rather than a tetrode. Self-oscillation in this stage is prevented by feeding back a small amount of energy from the plate circuit to the grid, this energy being out of phase (by π radians) with the energy being impressed on the grid from the preceding r-f stage. Such a method of preventing a tuned-plate, tuned-grid, r-f amplifier from developing oscillations is referred to as **neutralization.** The particular

circuit arrangement shown in Fig. 394 for accomplishing this is only one of several circuits which may be utilized to prevent self-oscillation.

In the transmitting circuit indicated in Fig. 394, the inductive reactances X_1 to X_4, inclusive, constitute what are known as h-f "chokes"; in each case they function to prevent a r-f "short circuit" of the grid. The capacitances, C_1, C_2, and C_3 form, in each case, a h-f path from the tuned plate circuit to the cathode.

The several anode circuits in both the audio and radio networks are supplied from well-filtered rectifier assemblies.

248. Radio-receiving Circuits. A great many different forms of circuits are utilized for the purpose of intercepting and demodulating

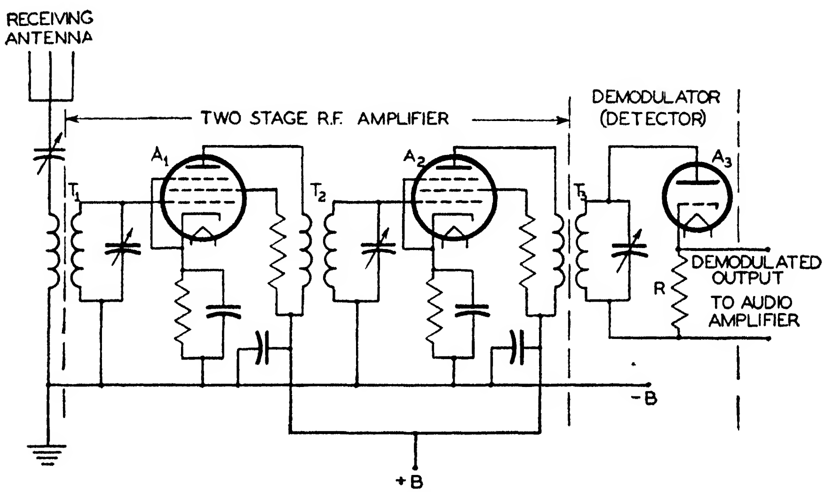


FIG. 396.—Schematic diagram of the circuit of a simple radiotelephone receiver.

electromagnetic waves. It will suffice for our purposes if we examine a simple receiving network that embodies only the essential elements required to recover the signal originally impressed on the carrier. Figure 396 is a diagrammatic sketch of a receiver circuit layout consisting of an absorbing, or antenna, system; a two-stage, r-f amplifier; and a thermionic half-wave rectifier A_3 acting as a demodulating member. It is assumed that the drop over the resistance R is being amplified by a suitable multistage audio amplifier, which in turn feeds a loud-speaker or other translating device. The potential that appears across R has the form shown in the bottom curve in Fig. 395.

As shown, the antenna can be tuned to the incident carrier wave by means of the series variable capacitor, though often (unfortunately) the antenna is arranged to function as an untuned absorbing unit.

The radio transformers T_2 and T_3 have tuned secondaries. The r-f amplifying tubes, A_1 and A_2, are pentodes, thus giving high amplification. Often more than two stages of r-f amplification are employed. The rotors of all tuning capacitors are usually mechanically connected ("ganged") so that they can be adjusted simultaneously by means of a single control member. Anode potential is supplied by a well-filtered, full-wave thermionic rectifying unit.

249. Guided-Wave (Carrier-Current) Communication. In 1910 the late General Squier (then Major) of the U.S. Army, by a bold and ingenious adaptation of the fundamental principles and apparatus previously employed in radio transmission, developed a new and highly important system of communication. Prior to General Squier's invention it had, of course, been possible to transmit simultaneously several

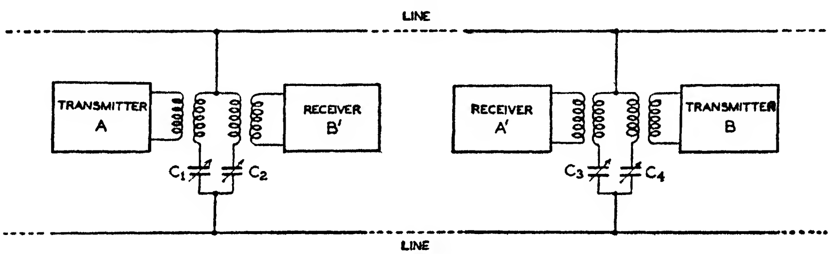


FIG. 397.—Carrier-current transmission and reception.

telegraph messages over a single circuit, and while certain earlier attempts had been made to develop a system of syntonic multiplex telegraphy, it is to General Squier that the world owes the production of a multiplex system of communication by which it is possible to carry on many simultaneous two-way telephone conversations over one electrical circuit. In other words, the possible traffic-carrying capacity of a simple metallic telephone or telegraph circuit has been increased many fold. How this is accomplished will be evident from the following description of the essentials of the Squier system.

The transmitting and receiving equipment utilized in the guided-wave system is essentially that employed in ordinary radio communication. However, instead of being connected to a radiating or absorbing system, consisting of an antenna and the ground, the transmitter and receiver are bridged across a physical telephone or telegraph pair, or between a single wire and the ground. By this arrangement, the energy is guided along the physical circuit in the form of an h-f alternating current instead of being radiated into space. This is shown diagrammatically in Fig. 397. The equipment there sketched would be required for a single two-way conversation. Transmitter A would be adjusted to deliver a modu-

lated carrier wave to the line, the frequency of the carrier current being, say, 60 kc. By means of the variable capacitance C_1, the line is then brought into resonance with *A*. Transmitter *B* would be set for some other noninterfering frequency, for instance, 40 kc. The receiving organization *A′* would be adjusted to resonance with transmitter *A*, and receiver *B′* is adjusted to resonance with transmitter *B*. Each party to the conversation would thus utilize a given carrier wave. At the same time that this "super-channel" is in operation, communication may also be maintained by ordinary telephonic or telegraphic means. Additional

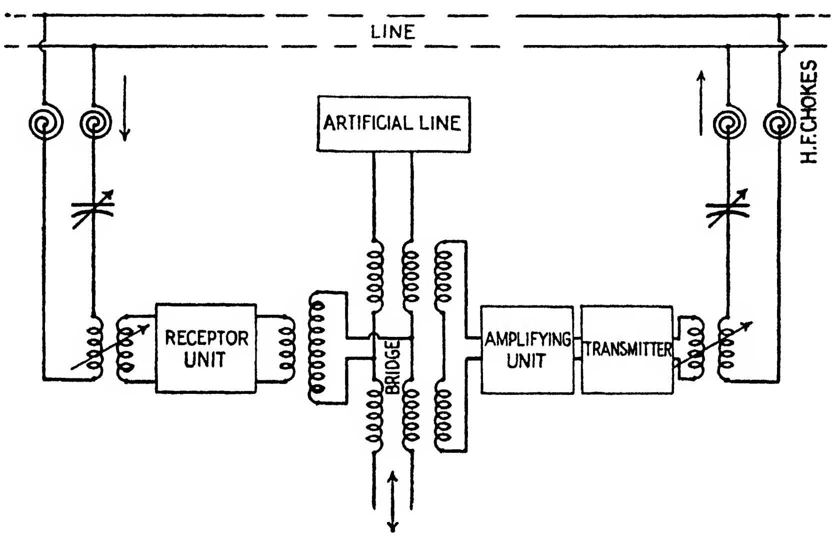


FIG. 398.—Circuit arrangement by means of which interaction between transmitting and receiving circuits is avoided in carrier-current transmission.

h-f channels may be established over the same physical pair, a definite pair of frequencies being assigned to each two-way channel. By using relatively high frequencies, and a coaxial cable instead of the usual two-wire circuit, it is now possible to provide for more than 200 guided-wave channels while utilizing only one physical circuit.

In practice it becomes necessary to provide means whereby interaction between adjacent apparatus, and also between interconnected circuits, may be eliminated. In Fig. 398 is shown how reaction between interconnected circuits is avoided. The bridge circuit (sometimes called a hybrid coil) should be carefully noted. By means of this arrangement of circuits the energy from the receptor is prevented from feeding into the transmitter and thus giving rise to regeneration and the consequent production of a-f oscillations ("howling"). The artificial line serves to

balance the constants of the local physical telephone circuit. When a multiplicity of two-way super channels is set up, the high frequency chokes are replaced by band filters (Sec. 167). A setup of circuits similar to that shown in the figure is required for each two-way h-f telephonic or telegraphic channel. Guided-wave (carrier-current) multiplex telephony and telegraphy are in extensive commercial use in this country and abroad.

The guided-wave system is employed to some extent as a means of communication over power transmission lines. When utilized for this purpose the transmitting and receiving circuits are electrostatically coupled to the power wires. Coupling is effected in one of two ways. In some installations one or more wires several hundred feet in length are supported parallel to and slightly below the power wires. In other plants the carrier current equipment is coupled to the power wire system through special high-potential condensers.

Recently considerable attention has been given to a method of controlling the propagation of electromagnetic waves by means of what are called **wave guides.** Such a "conductor" consists of a long metallic box that is circular or rectangular in cross section, the cross-sectional dimension being of the order of 2 or 3 cm. Microwaves generated by suitable means, at one end of such an enclosure, will be propagated down the guide as a result of repeated reflections from the side walls. It has been found that the attenuation is less than when using a standard transmission line or a coaxial cable. Waves of several different frequencies may be transmitted simultaneously through a given guide of this character. It is possible that such wave guides may come to be used for multiplex telephony and telegraphy, at least under certain circumstances. The reader will find an authoritative but simple discussion of such wave guides in a book by J. B. Hoag entitled "Basic Radio," pages 330 *ff*.

250. Radar. Radar, or radiolocation, may be said to be the science of the use of electromagnetic waves for the detection and location of an object, either fixed or moving. It is based on two fundamental principles: (1) that such waves can be concentrated into beams by means of suitable reflectors, and (2) that at least partial reflection of these waves will take place at the bounding surface between two mediums whose electrical properties differ. These principles were established by the original classical experiments of Hertz (Sec. 243). The determination of the distance of the reflecting body from the source of radiation also involves an old technique. The method used is an adaption of Fizeau's classical experiment to determine the velocity of light. The process by which airplanes, ships, and land masses may be located and their distance determined may be briefly outlined as follows.

In Fig. 399, S represents an ultra h-f transmitter (order of 1,000 mc) whose radiation is concentrated into a beam by means of a suitable wave reflector, consisting of a metallic mirror or wire grid. Equally spaced groups of waves are sent out as the beam is made to sweep over a certain

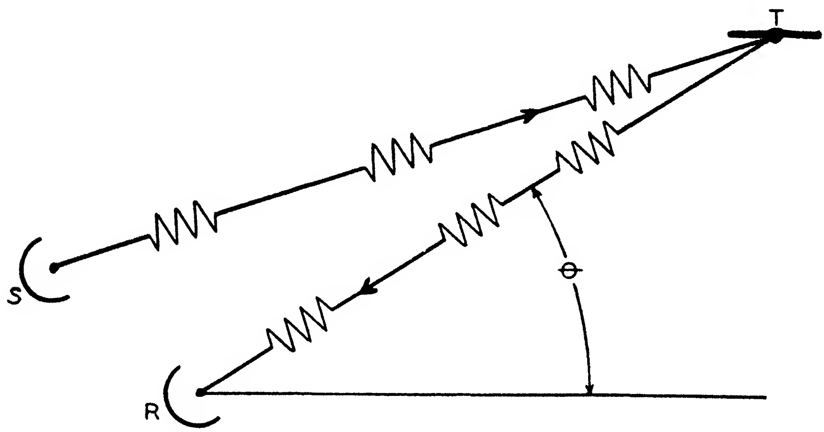


FIG. 399.—Illustrating the principle involved in radar.

area, in much the same way a search-light is operated. If and when this beam strikes a target, a limited amount of reradiation will occur. Some of this secondary radiation will reach the receiver (consisting of a dipole at the focus of the receiving mirror) and can be amplified. The amplified energy can be made to operate a timing device. It is this timing device which is, in some respects, unique.

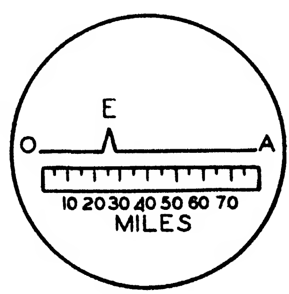


FIG. 400.—Appearance on the cathode-ray-tube screen of a reflected radar pulse.

It consists of a cathode-ray oscillograph so connected that, as a pulse of waves is emitted by the transmitter, the electronic beam is started on its sweep movement. The amplified returning wave (echo) pulses are applied to the vertical deflecting plates of the cathode-ray tube. If the sweep circuit is synchronized with the outgoing pulses, the received pulses (if the target is at a fixed distance) will appear superimposed on one another as vertical displacements from the horizontal time axis. The general character of the indication appearing on the fluorescent screen of the oscillograph is shown diagrammatically in Fig. 400. The line OA represents the time base, the distance OE indicating the time taken by an emitted pulse to traverse the distance from the transmitter to the target

and return. Knowing the velocity of electromagnetic waves to be 186,000 miles/sec and having at hand data concerning the sweep frequency, it is possible to calibrate the time base in miles. If a target were 25 miles away as shown in the figure, the distance OE on the time axis would represent 0.000269 sec. If the target is moving "in the line of sight" the pulse record on the screen will move toward or away from the point O. By increasing the frequency of the sweep circuit one can measure time intervals of the order of a few microseconds. Thus it becomes possible to determine accurately the location and distance of objects which may be relatively near, but which may not be visible. In this way, by the aid of portable radar units in airplanes, specific targets can be identified. Thus the equipment can be made to function as an altimeter.

When utilized for military purposes, such as the spotting of airplanes from a ground station, it of course becomes necessary to determine the angle (θ) made by the echo beam with the horizontal (altitude) and also its bearing, or azimuth. These observations are made by means of angle-measuring devices attached to the reflector assemblies.

By causing the cathode-ray beam to scan rapidly the entire fluorescent screen at the same time that the radio search beam scans a given target area, a maplike image of the target area will appear on the screen. Such a procedure can be carried out from great heights and when clouds or darkness make direct visual observation impossible. Thus we have a form of television based on the use of radio waves rather than on ordinary light. Both marine and aerial navigation will undoubtedly make wide use of radar equipment.

251. Facsimile Transmission. One of the important applications of the principles discussed in previous chapters is the transmission of pictures by wire and by radiotelegraphy. It is now possible to transmit photographs, including facsimiles of checks, drafts, legal documents, and the page of a newspaper between points separated by thousands of miles. Figure 401 is a reproduction of a photograph transmitted by this means. Several forms of equipment have been worked out for accomplishing this type of transmission, but the basic units in each case are the photoelectric cell (Sec. 186) and the thermionic tube used as an amplifier (Sec. 236). In Fig. 402 are shown the essential elements employed in a typical system for the transmission of pictures.

The transmitting equipment consists of a motor-driven cylinder C about which is wrapped the photograph, printed matter, or other copy to be telegraphed. The cylinder is caused to advance longitudinally as it revolves. Light from a suitable source S is focused, by means of the lens L, to a small point on the revolving picture. The light reflected

from the copy is collected by the parabolic mirror M and directed by the plane mirror M' into the photoelectric cell. As the pencil of light explores the surface of the picture a variable amount of light will be reflected, the quantity of reflected energy depending upon the degree of

FIG. 401.—Photograph transmitted over a wire telegraph circuit. (*Courtesy of Underwood and Underwood.*)

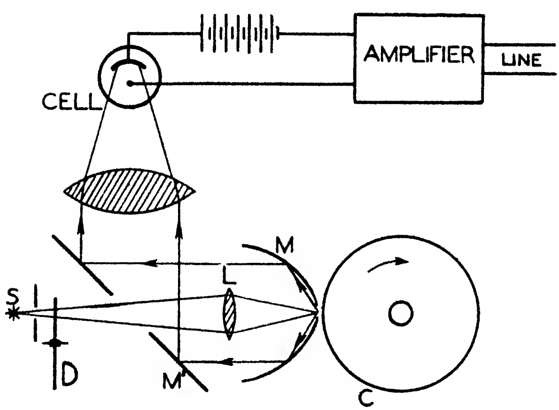


FIG. 402.—Diagrammatic sketch of a facsimile transmitter assembly.

coloring of the copy. The variation in intensity of the reflected beam of light thus incident on the sensitive surface of the photoelectric cell results in a corresponding variation in the photoelectric current. Amplification of this photoelectric current is necessary. This may be accomplished by

means of a d-c amplifier, or by the use of an ordinary audio amplifier operated in conjunction with a so-called "chopper." This is the plan sketched in Fig. 402. In order to make the potential supplied to the amplifier alternating or variable in character, a rotating disk D, having a row of apertures near its periphery, is positioned between the light source and the lens system. This "chopper" serves to interrupt the beam of light at a frequency of the order of 3,000 times per second, thus causing the output of the light cell to be of an intermittent character. The variations in the light incident on the cell, due to the scanning process, act to vary the amplitude of these regular variations imposed by the chopper. Various other plans are used for accomplishing this purpose.

The output of the photoelectric amplifier system may be made to modulate the carrier wave of a radio transmitter, the carrier current of a

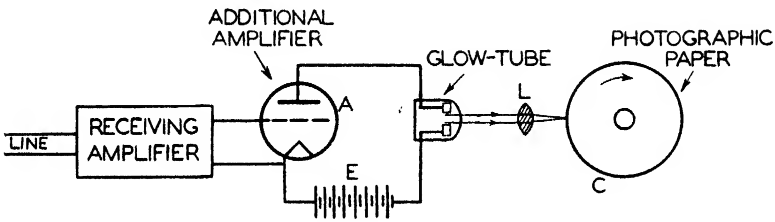


FIG. 403.—Showing the method by which facsimile reception is accomplished.

guided-wave circuit, or produce signals that may be transmitted over an ordinary physical telephone or telegraph line.

The reception of a picture or other copy is effected in a comparatively simple manner, one common method being indicated in Fig. 403. At the receiving terminal of the channel the picture-modulated carrier wave is passed into a standard receiving amplifier (to compensate for the attenuation which occurs during transmission) and thence to a special one-stage amplifier in the plate circuit of which is a glow tube. This glow tube is a simple two-electrode unit containing a gas such as argon or helium at low pressure. A current will pass through such a tube when a potential difference of from 200 to 400 volts is applied to its terminals, the gas being rendered incandescent by the passage of the current. The variable voltage due to the amplified signal being in series with the fixed potential difference supplied by E will give rise to a variable current through the glow tube. The intensity of the light emitted by such a unit is proportional to the strength of the current through the tube. Variations in the voltage impressed on the grid of the amplifier tube A due to the picture signals thus give rise to corresponding variations in the intensity of the light emitted by the glow lamp. This variable light

is focused by means of a suitable optical system on a piece of photographic paper which is wrapped about a revolving cylinder C; thus the gradations in light and shade of the original copy are impressed on the light-sensitive receiving sheet. The photographic paper later undergoes chemical development in the usual manner.

In carrying out facsimile transmission it is necessary that the receiving and transmitting cylinders revolve at synchronous speed. This may be accomplished by any one of several methods that need not be described here.

Facsimile picture transmission has been developed to the point where a 5- by 8-in. picture, or other similar copy, can be transmitted in about

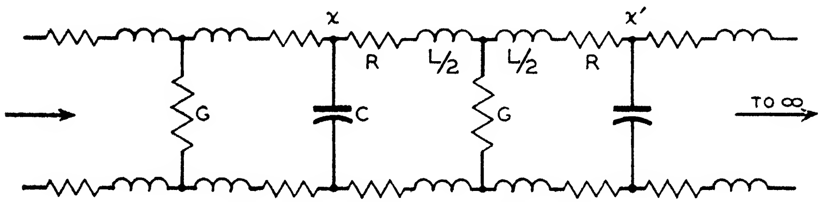


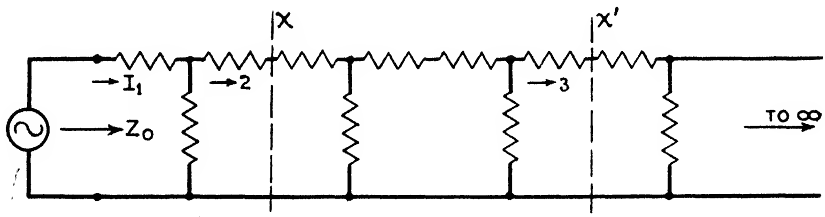
FIG. 404.—Infinite transmission line.

one minute. This is equivalent to a rate in excess of 600 words per minute, in the case of a printed or typed page. It seems probable that this form of transmission will in time replace the present code method of telegraphy. It is also probable that automatically recorded daily news bulletins will come to be distributed to offices and homes by means of facsimile transmission.

252. Transmission Lines. In view of the important part played by transmission lines in all branches of communication engineering, we may well conclude our study of electromagnetic waves by a consideration of the factors which have to do with the propagation of such waves along wire circuits.

The portion of the line between x and x' in Fig. 404 represents a very short section made up of ohmic resistances R, inductive and capacitive elements L and C, together with a certain amount of conductance G due to leakage between the conductors comprising the line. The line, as a whole, may be considered as extending to infinity. Physically, of course, there is no such thing as an infinite line. It is, however, convenient to assume that such a line may exist electrically. Such a line may be represented by a large number of small T sections connected in tandem, as shown in Fig. 405a. For our purposes we may consider one section only, as sketched in Fig. 405b. In that diagram, Z_1 is the total series impedance per section, Z_2 the total shunt impedance per section, Z_o

the characteristic impedance of the network, and Z_o' a load impedance. By **characteristic impedance** is meant the input impedance of one or more sections of the line as measured at the input end. If a line of finite length is terminated by an impedance Z_o' equal to the characteristic impedance, the line will behave **electrically** as if it were infinite in length. Under these conditions, all of the energy supplied to the line will be dissipated in the line and in the load impedance Z_o'; **no reflection will take**



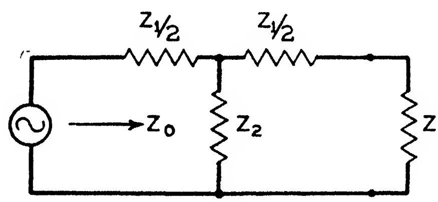


FIG. 405.—Equivalent of the line shown in Fig. 404.

place. Such a line is said to be **nonresonant.** It may be shown that

$$Z_o = \sqrt{Z_1Z_2 + \frac{Z_1^2}{4}}. \tag{290}$$

The characteristic impedance may also be expressed in terms of the line constants, thus,

$$Z_o = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{ohms,} \tag{291}$$

or

$$|Z_o| = \sqrt[4]{\frac{R^2 + \omega^2L^2}{G^2 + \omega^2C^2}} \quad \text{ohms.} \tag{292}$$

From Eq. (292) it is evident that the value of Z_o, for any given frequency, may be computed if the line constants are known. At very high frequencies, G is negligible and ωL becomes very large in comparison with R. Under these conditions, Z_o becomes practically a pure resistance whose

value is given by the expression

$$Z_o = \sqrt{\frac{L}{C}}, \tag{293}$$

and is thus seen to be independent of frequency. In general, the amplitude of the potential and current waves decrease as the waves move along the line toward the far end. However, the **ratio** of the potential difference across the line at any point, such as x in Fig. 405a, to the current at the same point is the same as the corresponding ratio **at any other point,** say at x'. This ratio gives the characteristic impedance Z_o of the line, and incidently provides us with another definition of this important line constant.

A knowledge of the impedance offered by a line is important for several reasons. The magnitude of the current at the input end of the line will be determined by the applied voltage and the characteristic impedance of the line, thus, $I_1 = V_1/Z_o$. Furthermore, the percentage of the energy that ultimately reaches the receiving end and is there absorbed by the load circuit depends upon the relation of the load impedance to the line impedance. It may be shown[1] that to attain a maximum transfer of power from the generator to a receiving network the impedance of the receiving circuit should equal the impedance of the generator. It is also true that to bring about a maximum transfer of energy from a transmission network to a load circuit the impedance of the load circuit should "match" that of the line. Thus it becomes necessary to compute, and to adjust, the impedance of the line in order that a maximum amount of energy may be transferred from the generator to the load circuit.

By the aid of Eq. (292) it is possible to develop working relations for the more common types of transmission lines. In the case of a pair of parallel wires for use at radio frequencies, the relation takes the form

$$Z_o \doteq 276 \log_{10} \frac{2S}{d} \qquad \text{ohms}, \tag{294}$$

where S is the distance between the wires and d the diameter of the wires.

For coaxial cables operated at high radio frequencies the equation is

$$Z_o \doteq 138 \log_{10} \frac{d_o}{d_i} \qquad \text{ohms}, \tag{295}$$

where d_o is the outside diameter of the inner cylinder and d_i the inside diameter of the outer cylinder. Coaxial cables have a Z_o value in the region of 77 ohms.

(It is suggested that the student consider this question: Does the

[1] See "Communication Engineering," 2d ed., p. 49, by W. L. Everitt.

length of a transmission line have any bearing on the magnitude of its characteristic impedance?)

In discussing the behavior of nonresonant lines, at least one other factor should be taken into account, viz., **attenuation,** *i.e.*, the decrease of the amplitude of the wave as it moves toward the receiving end of the line. It has been found that the decrease in the current value from section to section may be represented by the expression

$$p = \log_\epsilon \frac{I_1}{I_2}, \tag{296}$$

or

$$\frac{I_1}{I_2} = \epsilon^p, \tag{297}$$

where I_1 and I_2 represent the currents indicated in Fig. 405*a*, and p a factor known as the **propagation constant** per unit length or per section. Actually the constant p is a complex quantity, but for our purposes we will consider only the real part, which is designated as the **attenuation constant.** The last equation indicates that the attenuation is logarithmic and can therefore be expressed in decibels per unit length of line. Expressed in decibels the line attenuation will be

$$\text{Loss in db} = 20 \log_{10} \frac{I_s}{I_r}, \tag{298}$$

where I_s is the magnitude of the current entering the line and I_r the current leaving a given section of the line. For a No. 19 cable operated at 1,000 cycles the line loss is about 1 db per loop mile. Coaxial cable has a lower attenuation factor than open-line wires. When long transmission lines are operated at audio frequencies, the attenuations is of such magnitude that it becomes necessary to install amplifiers at intervals of 30 to 50 miles along the line.

The attenuation constant is a function of the frequency. It thus becomes evident that the components of a complex wave will not undergo equal attentuation; distortion will therefore result. This raises the question as to whether it is possible to arrange conditions so that all audio frequencies will undergo the same percentage attenuation, and thus preserve the quality characteristics of the signal being transmitted.

It may be shown[1] that distortionless transmission will obtain when

$$Z_o = \sqrt{\frac{R}{G}} = \sqrt{\frac{L}{C}}. \tag{299}$$

[1] See "Communication Engineering," pp. 116 ff., by W. L. Everitt, or "Communication Circuits," p. 64, by Ware and Reed.

In practice it is difficult, if not impossible, to attain the condition where $R/G = L/C$. Since, in the case of well-built lines, G is small, in order to reduce distortion to a minimum, it becomes necessary to make L large or R very small. If G is increased the attenuation will be increased; hence such a step is not feasible. Engineering economics places a limit on the maximum size of the conductor, and therefore upon the extent to which one can reduce R. In 1900 Professor Pupin suggested that the line inductance be increased by the introduction of **loading coils** at frequent intervals along a telephone line. By increasing the L factor in this way, long-distance transmission was greatly improved. That the transmission characteristics of a line may be materially improved is

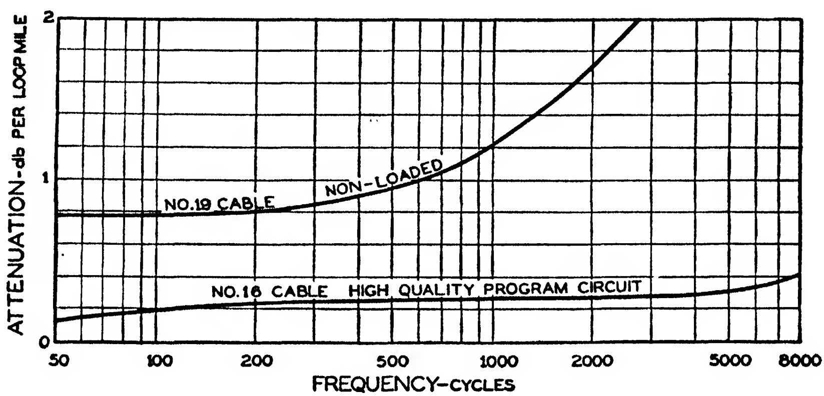


FIG. 406.—Showing comparative attenuation in the case of loaded and nonloaded transmission lines.

indicated by the graphs shown in Fig. 406. In recent years, however, carrier current transmission is replacing loaded-line operation.

Thus far in our discussion of transmission lines we have considered only the infinite line, or its electrical equivalent. If a line **does not** terminate in an impedance corresponding to the characteristic impedance, **wave reflection will take place.** The percentage of energy reflected will depend upon the degree of impedance mismatch that obtains. If the impedance ratio differs from unity, some reflection will occur. If the line is open, or if it is shorted, all of the energy reaching the far end of the line will be reflected; and under suitable conditions **standing waves** will be produced.

Suppose we have a situation such as that sketched in Fig. 407*a*. A pair of relatively long wires (W, W') spaced a few centimeters apart are coupled inductively to an oscillator that is capable of supplying a few watts of r-f energy. B is a conducting movable bar. Arrangements are

provided whereby a hot-wire or thermocouple type of milliammeter may be moved along the system while its terminals are in contact with the wires. First the oscillator is adjusted to give a frequency of such a magnitude that the wave length will be a fraction of the length of the wires W and W'. Then, if the bridging meter is set at some position such as A, and the shorting bar B is moved along the wires, it will be found that for some position of B the meter will read a maximum. If now we keep B in a fixed position and move the meter along the system, points will be found where maximum and minimum readings occur. This means that there are points where no potential difference exists between the wires; this in turn indicates that, due to interference, standing potential waves have been established on the wires which constitute the system. The line is now in resonance with the driver. Such points as n and n'

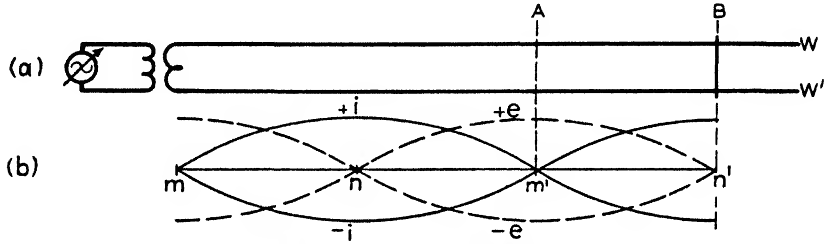


FIG. 407.—Standing waves on parallel wires a Lecher system.

are potential nodes; m and m' are current nodes. As in the corresponding case for an organ pipe, the distance nn' or mm' is equivalent to a **half wave length.** Thus we have a means whereby the wave length, and hence the frequency, of a given oscillator may be readily determined. The method is particularly useful when dealing with oscillations whose wave length is expressed in centimeters. This plan of wave length determination was originally suggested by Lecher, and such an assembly is referred to as a **Lecher system.**

In the case of a pair of narrowly spaced wires, or a coaxial cable, the loss of energy by radiation is negligible. Because of this important property both resonant and nonresonant lines may be, and are, widely used for the purpose of connecting a radio transmitter with its associated antenna system. When a nonresonant line is utilized for such a purpose provision must be made for matching both the transmitting circuit and the radiating system to the line impedance. In h-f transmission, where a dipole is used as a radiating system, a resonant line of the quarter-wave type is frequently employed. Details concerning the use of transmission lines for the purpose of conveying h-f energy from a generator to a radiating system will be found in any standard work on radio engineering.

253. What of the Future? By means of electromagnetic waves, guided or propagated through space, communication is now possible to all parts of the world. We speak to our neighbor in the next block or to some one halfway around the world. Distant scenes are made visible. Certain atoms have been made to yield vast quantities of energy, thereby changing the course of history. X rays can be made to penetrate several feet of metal, thus revealing its internal structure. With the aid of the electron microscope man is able to probe more deeply into nature's hidden secrets. And all these advances have come about through the research work of those pioneers of scientific progress whose contributions we have studied in the course now drawing to a close.

Is there much of consequence yet undiscovered? The answer is an emphatic, Yes! As a matter of fact, we have made only a beginning. The efforts of the analytical physicist and the research worker will continue to lead us into new and vast areas of discovery. The organizing and developmental skill of the engineer will make these new discoveries useful to all people of the earth. The science of electricity and magnetism has brought countless blessings to mankind. In the years to come it will make the earth a still better place in which to live and have our being.

INDEX

A

B

C